

(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2002 年 11 月 14 日 (14.11.2002)

PCT

(10) 国際公開番号
WO 02/091356 A1

(51) 国際特許分類: G10L 15/00
(21) 国際出願番号: PCT/JP02/04378
(22) 国際出願日: 2002 年 5 月 1 日 (01.05.2002)
(25) 国際出願の言語: 日本語
(26) 国際公開の言語: 日本語
(30) 優先権データ:
特願 2001-135423 2001 年 5 月 2 日 (02.05.2001) JP

(71) 出願人 (米国を除く全ての指定国について): ソニー株式会社 (SONY CORPORATION) [JP/JP]; 〒141-0001 東京都品川区北品川 6 丁目 7 番 3 5 号 Tokyo (JP).

(72) 発明者; および
(75) 発明者/出願人 (米国についてのみ): 廣江 厚夫 (HIROE, Atsuo) [JP/JP]; 〒141-0001 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社内 Tokyo (JP). 南野 活樹 (MINAMINO, Katsuki) [JP/JP]; 〒141-0001 東京

都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社内 Tokyo (JP). 河本 献太 (KAWAMOTO, Kenta) [JP/JP]; 〒141-0001 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社内 Tokyo (JP). 佐部 浩太郎 (SABE, Koh-taro) [JP/JP]; 〒141-0001 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社内 Tokyo (JP). 大橋 武史 (OHASHI, Takeshi) [JP/JP]; 〒141-0001 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社内 Tokyo (JP).

(74) 代理人: 小池 晃, 外 (KOIKE, Akira et al.); 〒105-0001 東京都港区虎ノ門二丁目 6 番 4 号 第 1 1 森ビル Tokyo (JP).

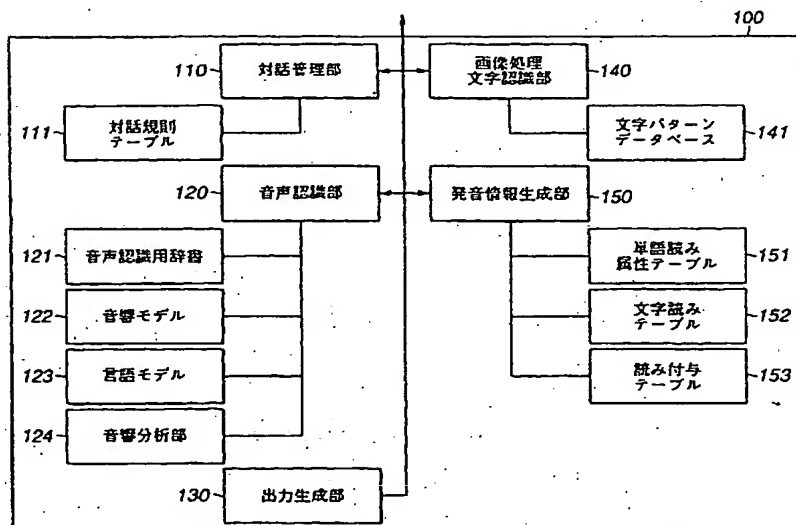
(81) 指定国 (国内): CN, KR, US.

添付公開書類:
— 国際調査報告書

2 文字コード及び他の略語については、定期発行される各 PCT ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

(54) Title: OBOT DEVICE, CHARACTER RECOGNIZING APPARATUS AND CHARACTER READING METHOD, AND CONTROL PROGRAM AND RECORDING MEDIUM

(54) 発明の名称: ロボット装置、並びに、文字認識装置及び文字認識方法、並びに、制御プログラム及び記録媒体



110...DIALOG MANAGEMENT UNIT.
111...DIALOG RULE TABLE
120...SPEECH RECOGNITION UNIT
121...DICTIONARY FOR SPEECH RECOGNITION
122...ACOUSTIC MODEL
123...LANGUAGE MODEL
124...ACOUSTIC ANALYSIS UNIT
130...OUTPUT GENERATION UNIT
140...IMAGE PROCESSING CHARACTER RECOGNITION UNIT
141...CHARACTER PATTERN DATABASE
150...PRONUNCIATION INFORMATION GENERATION UNIT
151...WORD READING ATTRIBUTE TABLE
152...CHARACTER READ TABLE
153...KANA IMPARTING TABLE

(57) Abstract: Characters estimated from the result of character recognition of an image picked up by a CCD camera (20), reading kana estimated from the characters, and pronunciation corresponding to the reading kana are generated by a pronunciation information generating unit (150), and one reading kana and pronunciation (how to read) is specifically selected from among candidates generated by matching the acquired readings with the utterance by the user collected by a microphone (23).

[続葉有]

WO 02/091356 A1

BEST AVAILABLE COPY



(57) 要約

CCDカメラ（20）において撮像された画像の文字認識の結果から推定される複数の文字と、これら各文字から推定される複数の読み仮名と、各読み仮名に対応する読み方とを発音情報生成部（150）において生成し、ここで得られた複数の読み方とマイク（23）において取得したユーザからの発声とをマッチングすることによって、生成された複数候補の中から1つの読み仮名及び発音のしかた（読み方）を特定する。

明細書

ロボット装置、並びに、文字認識装置及び文字認識方法、並びに、制御プログラム及び記録媒体

技術分野

本発明は、内部状態に応じて自律的に動作するロボット装置、並びに、文字認識装置及び文字認識方法、並びに、制御プログラム及び記録媒体に関し、特に、提示された文字とともに発音される音声に対して、撮像した画像から文字を認識し取得した音声をこの文字の発音として対応付けて新規に登録するロボット装置、並びに、提示された文字とともに発音される音声に対して、撮像した画像から文字を認識し取得した音声をこの文字の発音として新規に登録する文字認識装置及び文字認識方法、並びに、提示された文字とともに発音される音声に対して、撮像した画像から文字を認識し取得した音声をこの文字の発音として新規に登録する処理を実行させる制御プログラム及びこの制御プログラムが記録された記録媒体に関する。

背景技術

電氣的又は磁氣的な作用を用いて人間（生物）の動作に似た運動を行う機械装置を「ロボット」という。我が国においてロボットが普及し始めたのは、1960年代末からであるが、その多くは、工場における生産作業の自動化・無人化等を目的としたマニピュレータや搬送ロボット等の産業用ロボット（Industrial Robot）であった。

最近では、人間のパートナーとして生活を支援する、すなわち住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットの開発が進められている。このような実用ロボットは、産業用ロボットとは異なり、人間の生活環境の様々な局面において、個々に個性の相違した人間、又は様々な環境への

適応方法を自ら学習する能力を備えている。例えば、犬、猫のように4足歩行の動物の身体メカニズムやその動作を模した「ペット型」ロボット、或いは、2足直立歩行を行う動物の身体メカニズムや動作をモデルにしてデザインされた「人間型」又は「人間形」ロボット（Humanoid Robot）等の脚式移動ロボットは、既に実用化されつつある。これらの脚式移動ロボットは、動物や人間の容姿にできるだけ近い外観形状を有し、産業用ロボットと比較して動物や人間の動作に近い動作を行うことができ、更にエンターテインメント性を重視した様々な動作を行うことができるため、エンターテインメントロボットと呼称される場合もある。

脚式移動ロボットの中には、「目」に相当する小型カメラや、「耳」に相当する集音マイク等を備えているものもある。この場合、脚式移動ロボットは、取得した画像に対して画像処理を施すことによって、画像情報として入力した周囲の環境を認識したり、入力した周囲の音から「言語」を認識したりできる。

特に、外部から取得した音声を認識して文字に変換したり、音声を認識して応答したりする手法は、脚式移動ロボット以外にもパーソナルコンピュータや、その他の電子機器に音声認識装置として適用されている。

従来の音声認識の手法では、単語の発音と表記とが対応付けされて記憶された音声認識用の辞書（以下、認識用辞書と記す。）を用いて音声認識している。そのため、認識用辞書に登録されていない単語に関しては認識できないという欠点があった。更に、「文」のような連続した単語の発音を認識する場合には、認識用辞書に登録されている単語の組み合わせでなくてはならない。つまり、認識用辞書に登録されていない単語が含まれる場合、誤認識されるか、認識できない。

「北品川」という単語を例にとると、「北品川」が認証用辞書に登録されていなければ、「北品川」及び「北品川」を含む発音、例えば、「北品川は、どこですか。」という単語の連続からなる音声は、認識できないか、「北品川」の部分が誤認識される。そこで、認識用辞書に登録されていない単語を認識できるようにするためには、未登録の単語を新たに追加登録することが必要になる。

音声認識装置が音声認識を可能とするために備える認識用辞書とは、他の単語と区別するための識別子としての「単語シンボル」と、その単語の発音情報を表す「P L U列」とが対応付けられたものである。P L U（Phonone-like unit）と

は、音響的及び音韻的単位となるものである。発音された音声は、P L Uの組み合わせ（P L U列）として必ず表現することができる。

したがって、認識用辞書に単語を登録する場合は、単語シンボルとこれに対応するP L U列とを追加すればよい。ただし、単語シンボルとP L U列とを追加できる場合というのは、「北品川」や「k i t a s h i n a g a w a」という表記を、例えば、キーボード等のような入力手段を用いて直接入力できる場合に限られる。

そのため、ロボット装置のようにキーボードのような入力手段を備えていない場合には、音声として取得した単語の発音を音声認識して未知単語のP L U列を得る方法もある。この場合、ガーベージモデル（garbage model）を適用して認識している。ガーベージモデルとは、図20A及び図20Bに示すように、音声を発音の基本的な単位となる「音素」の組み合わせとして表した、また、単語の読み方の基本的な単位となる「かな」の組み合わせとして表した（ただし、日本語の場合。）モデルである。

従来の音声認識装置では、ガーベージモデルを適用することによって、音声による認識結果が得て、この認識結果に単語シンボルを当てはめて、これらに対応させて新規単語として認識用辞書に登録している。

ただし、ここで「音素」と「P L U」とは、ほぼ同義の単語として使用しており、「P L U列」は、複数の「P L U」が接続されることで構成された単語の発音を表記したものである。

ところが、ガーベージモデルを適用した従来の音声認識の手法では、同じ単語であってもユーザ毎に発声のしかたに微妙な違いがあることや、弱い音素（例えば、語頭の／s／等）は、必然的に認識されにくくなることや、周囲の雑音の影響による音素の変化や、音声区間検出の失敗等が原因となって、認識精度が悪くなるという欠点があった。

特に、ロボット装置に音声認識装置を適用した場合、音声認識装置側の音声取得用のマイクとユーザ（音声源）との距離が離れている状況下で使用されることが多いため、誤認識の頻度が高くなる。

具体的に、例えば、「きたしながわ」を認識させる場合について示すと、認識結果は、「h i t o t s u n a n o g a」や「i t a s n a g a :」

のように「きたしながわ」と類似しているが、同一ではないP L U列として認識されることがある。このような方法で単語登録された辞書を用いて音声認識を行うと、認識精度の低下、また誤認識による表示誤り等の問題が発生する。つまり、新規登録語には、不正確なP L U列が付与されていることになるため、この単語を認識する際の精度が低下するという問題点があった。

登録した人とは別の人が同じ単語を発音した場合、仮に「きたしながわ」が認識用辞書に登録されていたとしても、ユーザ毎の発音の癖から「きたしながわ」という単語を含む発音が認識されないこともあった。

また、音声認識の結果を文字に変換して表示する場合、新規登録語には、表示に関する情報が与えられていないため、誤った文字が表示されることがある。ユーザが「きたしながわ」を音声で登録した後、音声認識装置に対して「北品川に行きたい。」と発声した場合、音声認識装置には「きたしながわ」が正しく認識されたとしても、表示は「h i t o t s u n a n o g aに行きたい」や「『ひとつながの』に行きたい」になることがある。また、音声認識装置が認識結果のP L U列を音声合成で反復する場合も、合成された新規登録語のP L U列の部分だけが不自然な繋がりとして発声されるという不都合も生じる。

更に、このようにガーベージモデルによって登録された新規登録語は、品詞や意味等の単語の属性に関する情報を登録することができない。例えば、「北品川」を登録したとしても、この単語が名詞であるか地名であるかを表す情報を登録することができない。そのため、仮に、例えば、対話用の文法や認識用の言語モデル等に「<地名を表す語>+は+どこ+です+か」のような特定表現のための文法規則が予め記録されていたとしても、新規登録語には適用できないという問題点があった。登録時に単語の属性についても音声で入力することができるが、ユーザが単語の属性を知っている必要があった。また、単語の登録操作に加えて属性を入力することはユーザにとって煩わしい。

発明の開示

そこで本発明は、このような従来の実情に鑑みて提案されたものであり、提示

された文字とともに発音される音声に対して、撮像した画像から文字を認識し取得した音声をこの文字の発音として認識することによって、未登録の単語を新規単語として認識用辞書に登録でき、更に登録された新規単語を精度よく認識できるロボット装置、並びに、提示された文字とともに発音される音声に対して、撮像した画像から文字を認識し取得した音声をこの文字の発音として認識することによって、未登録の単語を新規単語として認識用辞書に登録でき、登録された新規単語を精度よく認識できる文字認識装置、及び、提示された文字を撮像し、撮像された画像から文字を認識し、提示とともに発音された音声を取得して認識された文字の発音として認識することによって、認識用辞書に新規単語として登録する文字認識方法、並びに、撮像した画像から文字を認識し取得した音声をこの文字の発音として新規に登録する処理を実行させる制御プログラム及びこの制御プログラムが記録された記録媒体を提供することを目的とする。

上述した目的を達成するために、本発明にかかるロボット装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から文字を抽出する文字認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、文字認識手段において抽出された文字から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

このようなロボット装置は、文字認識手段において抽出された文字から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成し、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音の

しかたであるとして音声認識用辞書に新規に記憶する。

また、本発明にかかる文字認識装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から文字を抽出する文字認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、文字認識手段において抽出された文字から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

このような文字認識装置は、文字認識手段において抽出された文字から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成し、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する。

また、本発明にかかる文字認識方法は、被写体を撮像する撮像工程と、撮像工程において撮像された画像から文字を抽出する文字認識工程と、周囲の音を取得する集音工程と、集音工程において取得された音から音声を認識する音声認識工程と、文字認識工程において抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成工程と、発音情報生成工程において生成された各音声波形と音声認識工程において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶

制御工程とを備える。

このような文字認識方法によれば、文字認識工程において抽出された文字から推定される複数通りの表音文字が単語表音テーブルに基づいて付与され、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形が生成され、発音情報生成工程において生成された各音声波形と音声認識工程において認識された音声の音声波形とが比較され、最も近い音声波形が抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶される。

更に、本発明にかかる制御プログラムは、被写体を撮像する撮像処理と、撮像処理によって撮像された画像から文字を抽出する文字認識処理と、周囲の音を取得する集音処理と、集音処理によって取得された音から音声を認識する音声認識処理と、文字認識処理によって抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、発音情報生成処理によって生成された各音声波形と音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理とをロボット装置に実行させる。また、上述の制御プログラムを記録媒体に記録して提供する。

図面の簡単な説明

図 1 は本発明の一構成例として示すロボット装置の外観を示す外観図である。

図 2 は本発明の一構成例として示すロボット装置の構成を示す構成図である。

図 3 は本発明の一構成例として示すロボット装置における画像音声認識部の構成を示す構成図である。

図 4 は本発明の一構成例として示すロボット装置の音声認識用辞書を説明する図である。

図 5 は本発明の一構成例として示すロボット装置の単語読み属性テーブルを説

明する図である。

図 6 は本発明の一構成例として示すロボット装置の文字読みテーブルを説明する図である。

図 7 は本発明の一構成例として示すロボット装置が新規単語を音声認識用辞書に登録する処理を説明するフローチャートである。

図 8 は本発明の一構成例として示すロボット装置の新規単語用認識用辞書を説明する図である。

図 9 は本発明の一構成例として示すロボット装置が認識した文字列の発音のしかた（読み方）を生成する処理を説明するフローチャートである。

図 10 は本発明の一構成例として示すロボット装置の制御プログラムのソフトウェア構成を示す構成図である。

図 11 は本発明の一構成例として示すロボット装置の制御プログラムのうち、ミドル・ウェア・レイヤの構成を示す構成図である。

図 12 は本発明の一構成例として示すロボット装置の制御プログラムのうち、アプリケーション・レイヤの構成を示す構成図である。

図 13 は本発明の一構成例として示すロボット装置の制御プログラムのうち、行動モデルライブラリの構成を示す構成図である。

図 14 は本発明の一構成例として示すロボット装置の行動を決定するためのアルゴリズムである有限確率オートマトンを説明する模式図である。

図 15 は本発明の一構成例として示すロボット装置の行動を決定するための状態遷移条件を表す図である。

図 16 は本発明の一構成例として示す人間型ロボット装置の前方からみた外観を説明する外観図である。

図 17 は本発明の一構成例として示す人間型ロボット装置の後方からみた外観を説明する外観図である。

図 18 は本発明の一構成例として示す人間型ロボット装置の自由度構成モデルを模式的に示す図である。

図 19 は本発明の一構成例として示す人間型ロボット装置の制御システム構成を説明する図である。

図20Aは、「音素」を基本単位とするガーページモデルを適用した従来の音声認識方法を示す模式図であり、図20Bは、「かな」を基本単位とするガーページモデルを適用した従来の音声認識方法を示す模式図である。

発明を実施するための最良の形態

本発明の一構成例として示すロボット装置は、内部状態に応じて自律動作するロボット装置である。このロボット装置は、少なくとも上肢と体幹部と下肢とを備え、上肢及び下肢、又は下肢のみを移動手段とする脚式移動ロボットである。脚式移動ロボットには、4足歩行の動物の身体メカニズムやその動きを模倣したベット型ロボットや、下肢のみを移動手段として使用する2足歩行の動物の身体メカニズムやその動きを模倣したロボット装置があるが、本実施の形態として示すロボット装置は、4足歩行タイプの脚式移動ロボットである。

このロボット装置は、住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットであり、内部状態（怒り、悲しみ、喜び、楽しみ等）に応じて行動できるほか、4足歩行の動物が行う基本的な動作を表出できるエンターテインメントロボットである。

このロボット装置は、特に、「犬」を模した形体であり、頭部、胴体部、上肢部、下肢部、尻尾部等を有している。各部の連結部分及び関節に相当する部位には、運動の自由度に応じた数のアクチュエータ及びポテンシオメータが備えられており、制御部の制御によって目標とする動作を表出できる。

ロボット装置は、周囲の状況を画像データとして取得するための撮像部や、周囲の音声を取得するマイク部と、外部から受ける作用を検出するための各種センサ等を備えている。撮像部としては、小型のCCD（Charge-Coupled Device）カメラを使用する。

本実施の形態として示すロボット装置は、画像認識装置及び音声認識装置を備えており、CCDカメラにおいて撮像された画像から文字を抽出し、抽出された文字から推定される複数通りの読み仮名を付与し、付与された複数通りの読み仮名のそれぞれに相当する音声波形を生成する。そして、この音声波形とマイク部

において取得した音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかた（読み方）であるとして音声認識用辞書に新規に記憶することができるロボット装置である。

以下、本発明の一構成例として示すロボット装置について、図面を参照して説明する。

本実施の形態では、ロボット装置 1 は、図 1 に示すように、「犬」を模した形状のいわゆるペット型ロボットである。ロボット装置 1 は、胴体部ユニット 2 の前後左右に脚部ユニット 3 A、3 B、3 C、3 D が連結され、胴体部ユニット 2 の前端部に頭部ユニット 4 が連結され、後端部に尻尾部ユニット 5 が連結されて構成されている。

胴体部ユニット 2 には、図 2 に示すように、CPU (Central Processing Unit) 10、DRAM (Dynamic Random Access Memory) 11、フラッシュROM (Read Only Memory) 12、PC (Personal Computer) カードインターフェイス回路 13 及び信号処理回路 14 が内部バス 15 を介して相互に接続されることにより形成されたコントロール部 16 と、このロボット装置 1 の動力源としてのバッテリー 17 とが収納されている。また、胴体部ユニット 2 には、ロボット装置 1 の向きや動きの加速度を検出するための角速度センサ 18 及び加速度センサ 19 が収納されている。

頭部ユニット 4 には、外部の状況を撮像するための CCD (Charge Coupled Device) カメラ 20 と、使用者からの「撫でる」や「叩く」といった物理的な働きかけにより受けた圧力を検出するためのタッチセンサ 21 と、前方に位置する物体までの距離を測定するための距離センサ 22 と、外部音を集音するためのマイク 23 と、鳴き声等の音声を出力するためのスピーカ 24 と、ロボット装置 1 の「目」に相当する LED (Light Emitting Diode) (図示せず) 等が所定位置にそれぞれ配置されている。CCD カメラ 20 は、頭部ユニット 4 の向く方向にある被写体を所定の画角で撮像することができる。

各脚部ユニット 3 A～3 D の関節部分、各脚部ユニット 3 A～3 D と胴体部ユニット 2 との連結部分、頭部ユニット 4 と胴体部ユニット 2 との連結部分、尻尾部ユニット 5 と尻尾 5 A との連結部分には、自由度数分のアクチュエータ 25、

～25_n及びポテンシオメータ26₁～26_nがそれぞれ配設されている。アクチュエータ25₁～25_nは、例えば、サーボモータを構成として有している。サーボモータの駆動により、脚部ユニット3A～3Dが制御されて目標の姿勢、或いは動作に遷移する。

これら角速度センサ18、加速度センサ19、タッチセンサ21、距離センサ22、マイク23、スピーカ24及び各ポテンシオメータ26₁～26_n等の各種センサ並びにLED及び各アクチュエータ25₁～25_nは、それぞれ対応するハブ27₁～27_nを介してコントロール部16の信号処理回路14と接続され、CCDカメラ20及びバッテリー17は、それぞれ信号処理回路14と直接接続されている。

信号処理回路14は、上述の各センサから供給されるセンサデータや画像データ及び音声データを順次取り込み、これらをそれぞれ内部バス15を介してDRAM11内の所定位置に順次格納する。また信号処理回路14は、これとともにバッテリー17から供給されるバッテリー残量を表すバッテリー残量データを順次取り込み、これをDRAM11内の所定位置に格納する。

このようにしてDRAM11に格納された各センサデータ、画像データ、音声データ及びバッテリー残量データは、CPU10が当該ロボット装置1の動作制御を行う際に使用される。

CPU10は、ロボット装置1の電源が投入された初期時において、フラッシュROM12に格納された制御プログラムを読み出して、DRAM11に格納する。又は、CPU10は、図1に図示しない胴体部ユニット2のPCカードスロットに装着された半導体メモリ装置、例えば、いわゆるメモリカード28に格納された制御プログラムをPCカードインターフェイス回路13を介して読み出してDRAM11に格納する。

CPU10は、上述のように信号処理回路14よりDRAM11に順次格納される各センサデータ、画像データ、音声データ、及びバッテリー残量データに基づいて自己及び周囲の状況や、使用者からの指示及び働きかけの有無を判断している。

更に、CPU10は、この判断結果とDRAM11に格納した制御プログラム

とに基づき行動を決定する。CPU 10は、当該決定結果に基づいてアクチュエータ $25_1 \sim 25_n$ の中から必要とするアクチュエータを駆動することによって、例えば、頭部ユニット4を上下左右に動かしたり、尻尾部ユニット5の尻尾を動かしたり、各脚部ユニット3A乃至3Dを駆動して歩行させたりする。また、CPU 10は、必要に応じて音声データを生成し、信号処理回路14を介してスピーカ24に供給する。また、CPU 10は、上述のLEDの点灯・消灯を指示する信号を生成し、LEDを点灯したり消灯したりする。

また、CPU 10は、上述のようにロボットを自律的に制御するほかに、後述する対話管理部110等からの要求に応じてロボットを動作させる。

これらの基本的な構成によって、ロボット装置1は、自己及び周囲の状況や、使用者からの指示及び働きかけに応じて自律的に行動する。

更に、ロボット装置1は、認識した発音と認識した文字との対応を新規登録語として音声認識用辞書に登録するための構成として、胴体部ユニット2のコントロール部16に、画像音声認識部100を備えている。画像音声認識部100は、図3に示すように、対話管理部110と、音声認識部120と、出力生成部130と、画像処理文字認識部140と、発音情報生成部150とを有している。音声認識用辞書とは、図4に示すように、他の単語と区別するための識別子としての「単語シンボル」と、この単語に対応する発音情報を表す「PLU列」とを記録したテーブルである。この辞書を参照することによって、単語の発音のしかた（読み方）、又は、発音に対応する単語の表記が抽出できる。

具体的に、対話管理部110は、マイク23から入力したユーザの発話、対話履歴等から入力した音声に対する応答を生成する。対話管理部110は、対話規則テーブル111に記憶された種々の対話規則に基づいて、入力した音声に対する応答パターンを生成する。

音声認識部120は、ユーザの発話を対話管理部110で処理できる形式、例えば、テキスト形式、構文解析、対話用フレーム等に変換する。音声認識部120は、具体的には、音声認識用辞書121、音響モデル122、言語モデル123、音響分析部124等からなる。音響分析部124では、認識に必要な特徴量の抽出が微少時間間隔で行われる。例えば、得られた音声信号のエネルギー、零交

差数、ピッチ、周波数特性、及びこれらの変化量等が抽出される。周波数分析には、線形予測分析 (L P C)、高速フーリエ変換 (F F T)、バンドパスフィルタ (B P F) 等が用いられる。

音声認識部 1 2 0 は、音響モデル 1 2 2 と言語モデル 1 2 3 とを用いて、音響分析部 1 2 4 で生成された特徴量系列に対応する単語系列を決定する。認識手法としては、例えば、隠れマルコフモデル (Hidden Markov Model : 以下、HMM と記す。) 等が用いられる。

HMM とは、状態遷移確率と確率密度関数とをもつ状態遷移モデルであり、状態を遷移しながら特徴量系列を出力する確率値を累積して尤度を決定する。その尤度の値を「スコア」として音声認識用辞書に記憶されている単語の発音のしかたと後述する画像処理文字認識部において認識された文字に対して付与される発音のしかたとのマッチングに使用する手法である。HMM の遷移確率及び確率密度関数等は、学習用データに基づく学習過程を通じて、予め学習して用意される値である。

音響モデルは、音素 (P L U)、音節、単語、フレーズ、文等、それぞれの単位毎に用意することができる。例えば、日本語の仮名『あ』・『い』・『う』・『え』・『お』・『か』・『き』…『ん』を単位とする音響モデルを用いた場合、これらを組み合わせて接続することによって、『はい』、『いいえ』、『おはよう』、『いまなんじですか』等の言葉が構成できる。音素とは、単語の発音情報を表すものであり、音響的及び音韻的単位である。本明細書では、音素と P L U (Phonone-like unit) とを区別しないで使用している。発音された音声は、音素 (P L U) の組み合わせ (P L U 列) として必ず表現することができる。

HMM によれば、このように構成された言葉とマイク 2 3 において取得した音声の特徴量系列との類似度をスコアとして計算することができる。音響モデルから「言葉」を構成するための情報として、言語モデル 1 2 3 と音声認識用辞書 1 2 1 とが利用される。音声認識用辞書 1 2 1 とは、認識対象となる各単語を構成するための音響モデル (ここでは、仮名の一文字『あ』、『い』、・・・等を示す。) の接続のしかたを対応テーブルとして示した辞書であり、言語モデル 1 2 3 とは、単語と単語との接続のしかたの規則を示したものである。

以下に示す例では、「単語」とは、認識処理の上で発音する際に、1つの纏まりとして扱う方が都合がよい単位のことを示しており、言語学的な単語とは必ずしも一致しない。例えば、以下の例では「北品川」を一単語として扱う場合があるが、これを「北」「品川」という2単語として扱っても構わない。更に、「北品川駅」や「北品川駅はどこですか」を発音する上での一単語として扱うこともできる。

また、本明細書では、「読み仮名」とは、漢字、英単語の読み方を表記したひらがな又はカタカナの意として用い、「発音のしかた」とは、読み仮名の実際の発音をローマ字、又はローマ字と記号とを使用して表記したものであり、言語学における「音素記号」に相当する。

例えば、『～時から、～時まで』という文を扱う場合について考える。この場合、まず、『0 (ぜろ)』『1 (いち)』・・・『24 (にじゅうよん)』という単語と、『時 (じ)』『から』『まで』という言葉のそれぞれに関して、音響モデル122を参照することによって、単語の接続のしかたが決定される。

次に、『(数字を表す単語)』『時』『から』『(数字を表す単語)』『時』『まで』という各単語を言語モデル123を参照することによって、文を構成するための各単語の接続のしかたが決定される。

この音声認識用辞書121と言語モデル123とを用いてHMMを適用することによって、『1時から2時まで』や『2時から5時まで』等の文と入力される特徴量系列との類似度がスコアとして計算できる。その中で最も高いスコアを有する単語系列からなる文を音声認識結果として出力する。

音声認識処理におけるスコアの計算は、音響モデル122によって与えられる音響的なスコアと、言語モデル123によって与えられる言語的なスコアとを総合評価して行われる場合もある。

言語的なスコアとは、例えば、連続するn個の単語間の遷移確率、又は連鎖確率に基づいて与えられるスコアである。遷移確率は、予め、大量のテキストから統計的に求められた値であり、ここでは、この遷移確率を「nグラム」と呼称する。

なお、言語モデルは、文法やnグラム中に単語を直接記述する以外にも、単語

のクラス（単語をある基準や属性にしたがって分類したもの）を記述する場合もある。

例えば、地名を表す単語を集め、それに<地名>というクラス名称を与えた場合に「<地名>+は+どこ+です+か」という文法を記述したり、nグラム中に「<地名>+は+どこ」の遷移確率を用意しておくこともできる。この場合、 $n = 3$ であり、正確には、遷移確率は、 $P(\text{<地名>} | \text{は、どこ})$ である。

出力生成部130は、対話管理部110が生成した応答パターンを実際の動作に変換する。例えば、対話管理部110が「首を左右に振る+『いいえ』と発声する」という応答パターンを生成した場合、出力生成部130は、これを受けて「首を左右に振る」に対応する動作パターンを生成しCPU10に送るとともに、「いいえ」に対応する音声波形を生成しスピーカ24から出力する。

画像処理文字認識部140は、CCDカメラ20で取り込んだ画像に含まれる文字列を文字パターンデータベース141に基づいて識別する。文字パターンデータベース141には、ひらがな、カタカナ、漢字、アルファベット、記号類、必要に応じて各国語の文字等の画像パターンが格納されている。画像処理文字認識部140は、CCDカメラ20からの入力画像と文字パターンデータベース141に格納されている画像パターンとの間でマッチングを行い、入力画像に含まれている文字列を認識する。

発音情報生成部150は、画像処理文字認識部140で認識された文字列に対応する発音情報、つまり文字列の読み仮名を生成し、更にその発音のしかた（読み方）を生成する。例えば、入力画像から「北品川」という文字列が認識された場合、「きたしながわ」という読み仮名を生成し、PLU列で「kitashinagawa」という発音のしかた（読み方）を生成する。

単語読み属性テーブル151は、図4に示すように、単語（文字列）と読み仮名と属性の組を記述したテーブルである。属性とは、「地名」、「名前」、「動物」のように単語のもつ意味を示している。

画像処理文字認識部140で認識された文字列がこのテーブルに含まれている場合は、このテーブルから読み仮名を抽出することで、読み仮名からその文字列の発音のしかた（読み方）を確定できる。単語読み属性テーブル151は、音声

認識用辞書 1 2 1 とは独立に用意する。

認識用辞書の語彙数には、認識速度や精度や処理上の都合で上限がある（例えば 6 万 5 5 3 6 語）が、単語読み属性テーブル 1 5 1 にはそれらの制限とは関係なく単語を記述することができる。この単語読み属性テーブル 1 5 1 は、他の言語資源から流用することも可能である。例えば、仮名漢字変換プログラムや形態素解析プログラム等で使用されている辞書等を流用することもできる。

文字読みテーブル 1 5 2 は、図 6 に示すように、文字と読み仮名との対応が記述されたテーブルである。記号やアルファベットや単漢字毎に読み仮名を記述しておく。使用可能な文字全てについて読み仮名を記述しておけば、任意の文字列に対して読み仮名から発音のしかた（読み方）を付与することができる。

読み付与テーブル 1 5 3 は、2 つのテーブルだけでは読み仮名が付与できない場合に読み仮名を付与するための規則や、読み仮名が特定できない場合に、これを特定するための規則が記述してある。例えば、音読み及び訓読みの統一、長音化に関する規則、連濁の規則、繰り返し記号に関する規則、英単語に読みを付与する規則がある。

具体的には、長音化に関する規則とは、「・・・おう」「・・・えい」等を「・・・おー」「・・・えー」等に変換する規則である。この規則によって、例えば、「とうきょう」は、「とーきょー」に変換される。連濁の規則とは、例えば、「品川口」の読みを「しながわ（品川）」と「くち（口）」との結合から生成する場合に、「くち」を濁らせて「ぐち」にする規則である。また、繰り返し記号に関する規則とは、「々・、・ゞ・ゝ・ゞ」等の繰り返し記号に対応して読み仮名を付ける規則である。更に、英単語に読み仮名を付与する規則とは、英単語の語末に“e”がある場合は、“e”自体は、発音しないかわりに前の母音を母音読みする等の規則である。例えば、“take”に「テーク」という読み仮名を付与する際に、“a”に対して「エー」という読み仮名を付与し、“ke”に対して、単に「ク」という読み仮名を付与する規則である。

次に、認識用辞書に新規単語を登録する際の処理を、図 7 を用いて具体的に説明する。

まず、ステップ S. 1 において、単語登録のための単語登録モードに移行する。

単語登録モードへの移行は、例えば、ロボット装置1は、ユーザが発する「登録モード」や「言葉を覚えて」等の言葉をトリガとして単語登録モードに移行する。このほかに、操作ボタンを設け、この操作ボタンが押されたときに単語登録モードへ移行するようにしてもよい。

ステップS2において、ロボット装置1は、ユーザに対して、登録したい単語の表記をロボット装置1のCCDカメラ20の前に提示する旨の指示及び／又は提示に加えてユーザが登録したい単語の読み方を発声する旨の指示を促す。

ユーザに対する指示は、ロボット装置1が音声によって指示してもよいし、また、図示しない表示部に指示内容を表示する場合でもよい。ここでは、「北品川」という単語を例として説明する。ユーザによって提示される文字は、漢字でも仮名でもローマ字表記でもPLU列でも構わない。具体的には、ロボット装置1は、「北品川」、「きたしながわ」、「キタシナガワ」、「kitashinagawa」等の何れの表記も認識できる。

ステップS3において、ロボット装置1は、文字提示のみであるか、文字提示とともに発話があったかを判断する。文字提示だけの場合は、ステップS4へ進み、文字提示とともに発話があった場合は、後述するステップS8へと進む。それ以外、すなわち、発声のみの場合は、従来と同様にガーベージモデルによる認識処理を行う。

はじめに、文字提示のみの場合について説明する。文字提示のみの場合、ステップS4において、ロボット装置1における画像処理文字認識部140は、CCDカメラ20において撮像された画像にどのような文字列が含まれているかを文字パターンデータベース141に基づいて、文字認識(OCR:Optical Character Recognition)する。ここで、画像処理文字認識部140は、文字認識結果の候補が1つに絞り込めない場合、複数の候補を残す。例えば、「北品川」という文字に対して「比品川」という認識結果が得られた場合は、「比品川」も残す。

続いて、ステップS5において、ロボット装置1における発音情報生成部150は、ステップS4での認識結果として得られた文字列に対して、文字列の発音のしかた(読み方)を生成する。発音を生成する際の詳細は、後述する。発音生成処理によって、文字列に対して発音のしかた(読み方)が付与される。認識さ

れた文字列が複数ある場合及び／又は１つの文字列に対して複数の発音のしかたが有り得る場合には、全ての発音パターンが適用される。

ステップＳ６において、ロボット装置１は、上述のように生成された文字列に対する発音のしかた（読み方）が正しいか否か、又は、複数の読み方のうちどれを採用すべきかをユーザに確認する。発音のしかた（読み方）が一通りのみの場合は、「読み方は、〇〇で正しいですか。」のように質問する。ユーザが「正しい」や「はい」等の応答を返した場合は、ステップＳ７に進む。

また、発音のしかた（読み方）が複数通りある場合は、それぞれについて「読み方は、〇〇ですか。」のように質問する。ユーザが「正しい」や「はい」等の応答を返した読み方を採用してステップＳ７に進む。

ユーザから「いいえ」等の応答を受けた場合、すなわち、正しい読み方が存在しない場合、ステップＳ２若しくはステップＳ４の処理まで戻る。

以上の処理によって、新規単語の読みを確定した後、ステップＳ７に進み、取得した文字列とこの文字列に対する発音のしかた（読み方）とを対応付けて新規単語として認識用辞書に登録する。新規単語を追加する際、図４に示す単語シンボル欄には、提示された文字の認識結果を使用する。この文字列に対応するＰＬＵ列欄には、ステップＳ６において確定した発音のしかた（読み方）が記述される。新規単語に登録した後、登録モードを終了する。その後、更新された認識用辞書を音声認識に反映させるための処理、例えば、音声認識プログラムの再起動等を行う。

一方、ステップＳ３において、ユーザが文字を提示するとともに表記した文字を発声した場合について説明する。文字提示とともに発話があった場合は、両者から得られる情報を協調的に使用することによってＰＬＵ列等の発音情報を精度よく生成することができる。

具体的には、文字認識の結果から推定される複数の文字と、これら各文字から推定される複数の読み仮名と、各読み仮名に対応する発音のしかた（読み方）とを生成する。このようにして得られた複数の発音のしかた（読み方）とマイク２３において取得したユーザからの発声とをマッチングすることによって、上述のように生成された複数候補の中から１つの読み仮名及び発音のしかた（読み方）

を特定する。

文字提示とともに発話があった場合、ステップS 8において、ロボット装置1における画像処理文字認識部140は、CCDカメラ20において撮像された画像から文字認識する。ここで、画像処理文字認識部140は、文字認識結果の候補が1つに絞り込めない場合、複数の候補を残す。

続いて、ステップS 9において、ロボット装置1における発音情報生成部150は、ステップS 8での認識結果として得られた文字列に対して、文字列の読み仮名を生成する。発音生成処理によって、文字列に対して発音のしかた（読み方）が付与される。認識された文字列が複数ある場合及び／又は1つの文字列に対して複数の読み方が可能な場合には、全ての発音パターンが適用される。

次に、ステップS 10において、文字列と発音のしかた（読み方）とから、一時的に仮の認識用辞書を生成する。この辞書を以下、新規単語用認識用辞書と記す。例えば、CCDカメラ20によって撮像された「北品川」という文字が画像処理文字認識部140において、「北品川」と「比晶川」の2通りに認識されたとする。音声情報生成部150は、「北品川」と「比晶川」に読み仮名を付与する。「北品川」には「きたしながわ」が付与され、「比晶川」には「ひしょうがわ」と「くらあきらがわ」の2通りが付与され、更に両者の発音のしかた（読み方）、すなわち、PLU列が生成される。この場合の新規単語用認識用辞書を図8に示す。

ステップS 11において、新規単語用認識用辞書を用いて、ユーザからの発声に対して音声認識を行う。ここでの音声認識は、連続音声認識ではなく、単語音声認識である。新規単語用認識用辞書が生成されるよりも前にユーザが発話している場合は、その発話を録音しておき、その録音音声に対して音声認識を行う。ステップS 11における音声認識とは、新規単語用認識用辞書に登録されている単語の中からユーザの発話と音響的に最も近い単語を探し出すことである。ただし、ステップS 11の処理では、単語シンボルが同一であっても、PLU列が異なる場合は別の単語とみなす。

図8では、ここに登録されている3単語（2つの「比晶川」は別単語とみなす）の中から、ユーザの発話である「きたしながわ」に最も近い単語を探し出すことである。結果として、単語シンボルとPLU列との組を1つに特定することがで

きる。

新規単語用認識用辞書の中から単語シンボルとP L U列との組が特定されたら、ステップS 7において、これを正規の音声認識用辞書1 2 1に登録する。新規単語を登録した後、登録モードを終了する。その後、更新された認識用辞書を音声認識に反映させるための処理、例えば、音声認識プログラムの再起動等を行う。

以上示した処理によって、ロボット装置1は、音声認識用辞書1 2 1に記憶されていない単語を新規単語として登録できる。

上述したステップS 5とステップS 9での文字列の発音のしかた（読み方）の生成に関して、図9を用いて詳細に説明する。

まず、ステップS 2 1において、画像処理文字認識部1 4 0によって認識された文字列が仮名文字だけで構成されているか否かを調べる。ただし、ここでの仮名文字とは、ひらがな・カタカナのほかに長音記号「ー」や繰り返し記号「々・…」等も含む。文字列が仮名文字だけで構成されている場合は、ステップS 2 2において、認識された仮名文字をその文字列の読み方とする。このとき、長音化等の発音を若干修正する場合もある。

一方、ステップS 2 1において、画像処理文字認識部1 4 0によって認識された文字列が仮名文字以外の文字を含んでいる場合、ステップS 2 3において、その文字列が単語読み属性テーブル1 5 1に含まれているか否かを判別する。

文字列が単語読み属性テーブル1 5 1に含まれている場合は、そのテーブルから読み仮名を取得し、更に発音のしかた（読み方）を生成する（ステップS 2 4）。また、単語読み属性テーブル1 5 1に単語の属性が記述されている場合は、属性も同時に取得する。この属性の利用方法については、後述する。

文字列が単語読み属性テーブル1 5 1に含まれていない場合、ステップS 2 5において、最長一致法・分割最小法、文字読みテーブル1 5 2に基づく読み付与、及び読み付与規則に基づく読み付与を組み合わせる読み仮名を取得する。

最長一致法・分割数最小法とは、単語読み属性テーブル1 5 1に含まれる単語を複数組み合わせることで入力文字列と同じものが構成できないか試みる方法である。例えば、入力文字列が「北品川駅前」である場合、これが単語読み属性テーブル1 5 1に含まれていなくても「北品川」と「駅前」とが含まれていれば、

これらの組み合わせから「北品川駅前」が構成できることから、結果として「きたしながわえきまえ」という読み方が取得できる。構成方法が複数通りある場合は、より長い単語が含まれる方を優先する（最長一致法）か、より少ない単語で構成できる方を優先する（分割数最小法）かして構成方法を選択する。

また、文字読みテーブル 1 5 2 に基づく読み付与とは、文字列を文字毎に分割し、分割した文字毎に文字読みテーブル 1 5 2 から読み仮名を取得する方法である。漢字の場合、1つの漢字には複数の読み仮名が付与できるため、文字列全体としての読み仮名は、各漢字の読み仮名の組み合わせになる。そのため、例えば、「音読みと訓読みとは混在しにくい」等の規則を用いて組み合わせの数を減らす方法である。

続いて、ステップ S 2 6 において、上述の各方法で取得したそれぞれの読み仮名の候補に対してスコア又は信頼度を計算し、高いものを選択する。これにより、入力された文字列に読み仮名を付与できる。得られた読み仮名から発音のしかた（読み方）を生成する。

ステップ S 2 2、ステップ S 2 4、ステップ S 2 6 のそれぞれの工程を経たのち、最終的に、ステップ S 2 7 において、読み仮名に対する発音のしかた（読み方）を長音化や連濁化等の規則に基づいて修正する。

ここで、単語読み属性テーブル 1 5 1 について詳細に説明する。音声認識用辞書 1 2 1 に単語を新規登録しただけでは、言語モデル 1 2 3 に記録された単語間の接続規則を適用することはできない。例えば、「北品川」を音声認識用辞書 1 2 1 に追加登録したとしても、それだけでは「北品川」に関する文法や「北品川」と他の単語との連鎖確率等は、生成されない。したがって、新規登録語に言語モデルの接続規則を反映させる方法は、理想的には、文法を追加したり、テキストデータから連鎖確率を計算し直したりして、言語モデルを構成し直すことであるが、以下に示す簡易的な方法によって新規登録後に言語モデルを適用することができる。

まず、言語モデルに含まれていない単語に＜未知語＞というクラス名を付ける。言語モデルには＜未知語＞と他の単語との連鎖確率を記述しておく。新規登録語は、＜未知語＞とみなし、この新規登録語と他の単語との連鎖確率は、＜未知語

>と他の単語との連鎖確率から計算する。

クラスとは、単語をある基準や属性にしたがって分類したものである。例えば、意味にしたがって分類し、それぞれを<地名>、<姓>、<国名>と命名したり、品詞にしたがって分類し、それぞれを<名詞>、<動詞>、<形容詞>と命名したりする。

言語モデルには、単語間の連鎖確率を記述するかわりにクラス間の連鎖確率やクラスと単語との連鎖確率を記述する。単語間の連鎖確率を求めるときは、単語がどのクラスに属するかを調べ、次に対応するクラスについての連鎖確率を求め、そこから単語間の連鎖確率を計算する。

新規登録語についても、どのクラスに属する単語であるかを登録時に推定することでクラスモデルが適用できる。

上述のようにすると未知語用モデルでは、新規登録語には、全て同一の値の連鎖確率が付される。それに対してクラスモデルでは、どのクラスに属するかによって異なる値になる。そのため一般的には、新規登録語についての言語的スコアは、クラスモデルを用いた方がより適切なスコアとなり、結果的に適切に認識される。

したがって、音声認識による単語登録において、従来、困難であったクラス名称が、容易に入力できる。すなわち、文字認識で得られた文字列（単語）が単語読み属性テーブル151に含まれている場合、このテーブルの属性欄からクラス名称を取得できる。なお、図5に示す例では、属性欄に属性を1つしか記述していないが、これを「<地名>、<固有名詞>、<駅名>」のように複数記述することもできる。この場合、例えば、<地名>というクラスが存在する場合は、<地名>、<固有名詞>、<駅名>の中から、クラス名称と一致する分類名、すなわち<地名>を採用する。

文字認識では、一文字ずつ認識するよりも、文字の連鎖に関する情報を含めて認識の方が精度が向上する場合がある。そこで、認識用辞書の「単語シンボル」欄や、単語読み属性テーブル151の「単語」欄等を文字の連鎖に関する情報として使用することによって、文字認識の精度を更に向上できる。

ところで、本実施の形態として示すロボット装置1は、内部状態に応じて自律

的に行動できるロボット装置である。ロボット装置 1 における制御プログラムのソフトウェア構成は、図 10 に示すようになる。この制御プログラムは、上述したように、予めフラッシュ ROM 12 に格納されており、ロボット装置 1 の電源投入初期時において読み出される。

図 10 において、デバイス・ドライバ・レイヤ 30 は、制御プログラムの最下位層に位置し、複数のデバイス・ドライバからなるデバイス・ドライバ・セット 31 から構成されている。この場合、各デバイス・ドライバは、CCD カメラ 20 (図 2) やタイマ等の通常のコンピュータで用いられるハードウェアに直接アクセスすることを許されたオブジェクトであり、対応するハードウェアからの割り込みを受けて処理を行う。

また、ロボティック・サーバ・オブジェクト 32 は、デバイス・ドライバ・レイヤ 30 の最下位層に位置し、例えば上述の各種センサやアクチュエータ 25₁ ~ 25_n 等のハードウェアにアクセスするためのインターフェイスを提供するソフトウェア群でなるバーチャル・ロボット 33 と、電源の切換え等を管理するソフトウェア群でなるパワーマネージャ 34 と、他の種々のデバイス・ドライバを管理するソフトウェア群でなるデバイス・ドライバ・マネージャ 35 と、ロボット装置 1 の機構を管理するソフトウェア群でなるデザインド・ロボット 36 とから構成されている。

マネージャ・オブジェクト 37 は、オブジェクト・マネージャ 38 及びサービス・マネージャ 39 から構成されている。オブジェクト・マネージャ 38 は、ロボティック・サーバ・オブジェクト 32、ミドル・ウェア・レイヤ 40、及びアプリケーション・レイヤ 41 に含まれる各ソフトウェア群の起動や終了を管理するソフトウェア群であり、サービス・マネージャ 39 は、メモリカード 28 (図 2) に格納されたコネクションファイルに記述されている各オブジェクト間の接続情報に基づいて各オブジェクトの接続を管理するソフトウェア群である。

ミドル・ウェア・レイヤ 40 は、ロボティック・サーバ・オブジェクト 32 の上位層に位置し、画像処理や音声処理等のこのロボット装置 1 の基本的な機能を提供するソフトウェア群から構成されている。また、アプリケーション・レイヤ 41 は、ミドル・ウェア・レイヤ 40 の上位層に位置し、当該ミドル・ウェア・

レイヤ40を構成する各ソフトウェア群によって処理された処理結果に基づいてロボット装置1の行動を決定するためのソフトウェア群から構成されている。

なお、ミドル・ウェア・レイヤ40及びアプリケーション・レイヤ41の具体的なソフトウェア構成をそれぞれ図11に示す。

ミドル・ウェア・レイヤ40は、図11に示すように、騒音検出用、温度検出用、明るさ検出用、音階認識用、距離検出用、姿勢検出用、タッチセンサ用、動き検出用及び色認識用の各信号処理モジュール50～58並びに入力セマンティクスコンバータモジュール59等を有する認識系60と、出力セマンティクスコンバータモジュール68並びに姿勢管理用、トラッキング用、モーション再生用、歩行用、転倒復帰用、LED点灯用及び音再生用の各信号処理モジュール61～67等を有する出力系69とから構成されている。

認識系60の各信号処理モジュール50～58は、ロボティック・サーバ・オブジェクト32のバーチャル・ロボット33によりDRAM11（図2）から読み出される各センサデータや画像データ及び音声データのうちの対応するデータを取り込み、当該データに基づいて所定の処理を施して、処理結果を入力セマンティクスコンバータモジュール59に与える。ここで、例えば、バーチャル・ロボット33は、所定の通信規約によって、信号の授受或いは変換をする部分として構成されている。

入力セマンティクスコンバータモジュール59は、これら各信号処理モジュール50～58から与えられる処理結果に基づいて、「うるさい」、「暑い」、「明るい」、「ボールを検出した」、「転倒を検出した」、「撫でられた」、「叩かれた」、「ドミソの音階が聞こえた」、「動く物体を検出した」又は「障害物を検出した」等の自己及び周囲の状況や、使用者からの指令及び働きかけを認識し、認識結果をアプリケーション・レイヤ41に出力する。

アプリケーション・レイヤ41は、図12に示すように、行動モデルライブラリ70、行動切換えモジュール71、学習モジュール72、感情モデル73及び本能モデル74の5つのモジュールから構成されている。

行動モデルライブラリ70には、図13に示すように、「バッテリー残量が少なくなった場合」、「転倒復帰する」、「障害物を回避する場合」、「感情を表現する場合」、

「ボールを検出した場合」等の予め選択されたいくつかの条件項目にそれぞれ対応させて、それぞれ独立した行動モデルが設けられている。

そして、これら行動モデルは、それぞれ入力セマンティクスコンバータモジュール59から認識結果が与えられたときや、最後の認識結果が与えられてから一定時間が経過したとき等に、必要に応じて後述のように感情モデル73に保持されている対応する情動のパラメータ値や、本能モデル74に保持されている対応する欲求のパラメータ値を参照しながら続く行動をそれぞれ決定し、決定結果を行動切換えモジュール71に出力する。

なお、この実施の形態の場合、各行動モデルは、次の行動を決定する手法として、図14に示すような1つのノード(状態) $NODE_0 \sim NODE_n$ から他のどのノード $NODE_0 \sim NODE_n$ に遷移するかを各ノード $NODE_0 \sim NODE_n$ に間を接続するアーク $ARC_1 \sim ARC_{n1}$ に対してそれぞれ設定された遷移確率 $P_1 \sim P_n$ に基づいて確率的に決定する有限確率オートマトンと呼ばれるアルゴリズムを用いる。

具体的に、各行動モデルは、それぞれ自己の行動モデルを形成するノード $NODE_0 \sim NODE_n$ にそれぞれ対応させて、これらノード $NODE_0 \sim NODE_n$ 毎に図15に示すような状態遷移表80を有している。

この状態遷移表80では、そのノード $NODE_0 \sim NODE_n$ において遷移条件とする入力イベント(認識結果)が「入力イベント名」の行に優先順に列記され、その遷移条件についての更なる条件が「データ名」及び「データ範囲」の行における対応する列に記述されている。

したがって、図15の状態遷移表80で表されるノード $NODE_{100}$ では、「ボールを検出(BALL)」という認識結果が与えられた場合に、当該認識結果とともに与えられるそのボールの「大きさ(SIZE)」が「0から1000」の範囲であることや、「障害物を検出(OBSTACLE)」という認識結果が与えられた場合に、当該認識結果とともに与えられるその障害物までの「距離(DISTANCE)」が「0から100」の範囲であることが他のノードに遷移するための条件となっている。

また、このノード $NODE_{100}$ では、認識結果の入力がない場合においても、

行動モデルが周期的に参照する感情モデル 7 3 及び本能モデル 7 4 にそれぞれ保持された各情動及び各欲求のパラメータ値のうち、感情モデル 7 3 に保持された「喜び (Joy)」、「驚き (Surprise)」若しくは「悲しみ (Sadness)」の何れかのパラメータ値が「50 から 100」の範囲であるときには他のノードに遷移することができるようになっている。

また、状態遷移表 8 0 では、「他のノードへの遷移確率」の欄における「遷移先ノード」の列にそのノード $NODE_0 \sim NODE_n$ から遷移できるノード名が列記されているとともに、「入力イベント名」、「データ名」及び「データの範囲」の行に記述された全ての条件が揃ったときに遷移できるほかの各ノード $NODE_0 \sim NODE_n$ への遷移確率が「他のノードへの遷移確率」の欄内の対応する箇所にそれぞれ記述され、そのノード $NODE_0 \sim NODE_n$ に遷移する際に出力すべき行動が「他のノードへの遷移確率」の欄における「出力行動」の行に記述されている。なお、「他のノードへの遷移確率」の欄における各行の確率の和は 100 [%] となっている。

したがって、図 1 5 の状態遷移表 8 0 で表されるノード $NODE_{100}$ では、例えば「ボールを検出 (BALL)」し、そのボールの「SIZE (大きさ)」が「0 から 1000」の範囲であるという認識結果が与えられた場合には、「30 [%]」の確率で「ノード $NODE_{120}$ (node 120)」に遷移でき、そのとき「ACTION 1」の行動が出力されることとなる。

各行動モデルは、それぞれこのような状態遷移表 8 0 として記述されたノード $NODE_0 \sim NODE_n$ が幾つも繋がるようにして構成されており、入力セマンティクスコンバータモジュール 5 9 から認識結果が与えられたとき等に、対応するノード $NODE_0 \sim NODE_n$ の状態遷移表を利用して確率的に次の行動を決定し、決定結果を行動切換えモジュール 7 1 に出力するようになされている。

図 1 2 に示す行動切換えモジュール 7 1 は、行動モデルライブラリ 7 0 の各行動モデルからそれぞれ出力される行動のうち、予め定められた優先順位の高い行動モデルから出力された行動を選択し、当該行動を実行すべき旨のコマンド（以下、これを行動コマンドという。）をミドル・ウェア・レイヤ 4 0 の出力セマンティクスコンバータモジュール 6 8 に送出する。なお、この実施の形態においては、

図 13 において下側に表記された行動モデルほど優先順位が高く設定されている。

また、行動切換えモジュール 71 は、行動完了後に出力セマンティクスコンバータモジュール 68 から与えられる行動完了情報に基づいて、その行動が完了したことを学習モジュール 72、感情モデル 73 及び本能モデル 74 に通知する。

一方、学習モジュール 72 は、入力セマンティクスコンバータモジュール 59 から与えられる認識結果のうち、「叩かれた」や「撫でられた」等、使用者からの働きかけとして受けた教示の認識結果を入力する。

そして、学習モジュール 72 は、この認識結果及び行動切換えモジュール 71 からの通知に基づいて、「叩かれた（叱られた）」ときにはその行動の発現確率を低下させ、「撫でられた（誉められた）」ときにはその行動の発現確率を上昇させるように、行動モデルライブラリ 70 における対応する行動モデルの対応する遷移確率を変更する。

他方、感情モデル 73 は、「喜び (Joy)」、「悲しみ (Sadness)」、「怒り (Anger)」、「驚き (Surprise)」、「嫌悪 (Disgust)」及び「恐れ (Fear)」の合計 6 つの情動について、各情動毎にその情動の強さを表すパラメータを保持している。そして、感情モデル 73 は、これら各情動のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール 59 から与えられる「叩かれた」及び「撫でられた」等の特定の認識結果と、経過時間及び行動切換えモジュール 71 からの通知と等に基づいて周期的に更新する。

具体的には、感情モデル 73 は、入力セマンティクスコンバータモジュール 59 から与えられる認識結果と、そのときのロボット装置 1 の行動と、前回更新してから経過時間と等に基づいて所定の演算式により算出されるそのときのその情動の変動量を $\Delta E[t]$ 、現在のその情動のパラメータ値を $E[t]$ 、その情動の感度を表す係数を k 。として、(1) 式によって次の周期におけるその情動のパラメータ値 $E[t+1]$ を算出し、これを現在のその情動のパラメータ値 $E[t]$ と置き換えるようにしてその情動のパラメータ値を更新する。また、感情モデル 73 は、これと同様にして全ての情動のパラメータ値を更新する。

$$E[t+1] = E[t] + k \times \Delta E[t] \quad \dots (1)$$

なお、各認識結果や出力セマンティクスコンバータモジュール 68 からの通知

が各情動のパラメータ値の変動量 $\Delta E[t]$ にどの程度の影響を与えるかは予め決められており、例えば「叩かれた」といった認識結果は「怒り」の情動のパラメータ値の変動量 $\Delta E[t]$ に大きな影響を与え、「撫でられた」といった認識結果は「喜び」の情動のパラメータ値の変動量 $\Delta E[t]$ に大きな影響を与えるようになっている。

ここで、出力セマンティクスコンバータモジュール68からの通知とは、いわゆる行動のフィードバック情報（行動完了情報）であり、行動の出現結果の情報であり、感情モデル73は、このような情報によっても感情を変化させる。これは、例えば、「吠える」といった行動により怒りの感情レベルが下がるといったようなことである。なお、出力セマンティクスコンバータモジュール68からの通知は、上述した学習モジュール72にも入力されており、学習モジュール72は、その通知に基づいて行動モデルの対応する遷移確率を変更する。

なお、行動結果のフィードバックは、行動切換えモジュレータ71の出力（感情が付加された行動）によりなされるものであってもよい。

一方、本能モデル74は、「運動欲 (exercise)」、「愛情欲 (affection)」、「食欲 (appetite)」及び「好奇心 (curiosity)」の互いに独立した4つの欲求について、これら欲求毎にその欲求の強さを表すパラメータを保持している。そして、本能モデル74は、これらの欲求のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール59から与えられる認識結果や、経過時間及び行動切換えモジュール71からの通知等に基づいて周期的に更新する。

具体的には、本能モデル74は、「運動欲」、「愛情欲」及び「好奇心」については、認識結果、経過時間及び出力セマンティクスコンバータモジュール68からの通知等に基づいて所定の演算式により算出されるそのときのその欲求の変動量を $\Delta I[k]$ 、現在のその欲求のパラメータ値を $I[k]$ 、その欲求の感度を表す係数 k_1 として、所定周期で(2)式を用いて次の周期におけるその欲求のパラメータ値 $I[k+1]$ を算出し、この演算結果を現在のその欲求のパラメータ値 $I[k]$ と置き換えるようにしてその欲求のパラメータ値を更新する。また、本能モデル74は、これと同様にして「食欲」を除く各欲求のパラメータ値を更新する。

$$I[k+1] = I[k] + k_i \times \Delta I[k] \quad \dots(2)$$

なお、認識結果及び出力セマンティクスコンバータモジュール68からの通知等が各欲求のパラメータ値の変動量 $\Delta I[k]$ にどの程度の影響を与えるかは予め決められており、例えば出力セマンティクスコンバータモジュール68からの通知は、「疲れ」のパラメータ値の変動量 $\Delta I[k]$ に大きな影響を与えるようになっている。

なお、本実施の形態においては、各情動及び各欲求（本能）のパラメータ値がそれぞれ0から100までの範囲で変動するように規制されており、また係数 k_e 、 k_i の値も各情動及び各欲求毎に個別に設定されている。

一方、ミドル・ウェア・レイヤ40の出力セマンティクスコンバータモジュール68は、図11に示すように、上述のようにしてアプリケーション・レイヤ41の行動切換えモジュール71から与えられる「前進」、「喜ぶ」、「鳴く」又は「トラッキング（ボールを追いかける）」といった抽象的な行動コマンドを出力系69の対応する信号処理モジュール61～67に与える。

そしてこれら信号処理モジュール61～67は、行動コマンドが与えられると当該行動コマンドに基づいて、その行動をするために対応するアクチュエータ25₁～25_n（図2）に与えるべきサーボ指令値や、スピーカ24（図2）から出力する音の音声データ及び又は「目」のLEDに与える駆動データを生成し、これらのデータをロボティック・サーバ・オブジェクト32のバーチャル・ロボット33及び信号処理回路14（図2）を順次介して対応するアクチュエータ25₁～25_n又はスピーカ24又はLEDに順次送出する。

このようにしてロボット装置1は、制御プログラムに基づいて、自己（内部）及び周囲（外部）の状況や、使用者からの指示及び働きかけに応じた自律的な行動ができる。したがって、上述した文字認識処理を実行するためプログラムを備えていないロボット装置に対しても、文字認識処理によって画像から抽出した文字の発音のしかたを音声認識処理によって周囲の音から認識された音声に基づいて決定する処理を実行するための制御プログラムを読み込ませることによって、図7に示した文字認識処理を実行させることができる。

このような制御プログラムは、ロボット装置が読取可能な形式で記録された記

録媒体を介して提供される。制御プログラムを記録する記録媒体としては、磁気読取方式の記録媒体（例えば、磁気テープ、フロッピー（登録商標）ディスク、磁気カード）、光学読取方式の記録媒体（例えば、CD-ROM、MO、CD-R、DVD）等が考えられる。記録媒体には、半導体メモリ（いわゆるメモリカード（矩形型、正方形型等形状は問わない。）、ICカード）等の記憶媒体も含まれる。また、制御プログラムは、いわゆるインターネット等を介して提供されてもよい。

これらの制御プログラムは、専用の読込ドライバ装置、又はパーソナルコンピュータ等を介して再生され、有線又は無線接続によってロボット装置1に伝送されて読み込まれる。また、ロボット装置は、半導体メモリ、又はICカード等の小型化された記憶媒体のドライブ装置を備える場合、これら記憶媒体から制御プログラムを直接読み込むこともできる。ロボット装置1では、メモリカード28から読み込むことができる。

なお、本発明は、上述した実施の形態のみに限定されるものではなく、本発明の要旨を逸脱しない範囲において種々の変更が可能であることは勿論である。本実施の形態では、4足歩行のロボット装置に関して説明したが、ロボット装置は、2足歩行であってもよく、更に、移動手段は、脚式移動方式に限定されない。

以下に、本発明の別の実施の形態として示す人間型ロボット装置の詳細について説明する。図16及び図17には、人間型ロボット装置200を前方及び後方の各々から眺望した様子を示している。更に、図18には、この人間型ロボット装置200が具備する関節自由度構成を模式的に示している。

図16に示すように、人間型ロボット装置200は、脚式移動を行う左右2足の下肢201R、201Lと、体幹部202と、左右の上肢203R、203Lと、頭部204とで構成される。

左右各々の下肢201R、201Lは、大腿部205R、205Lと、膝関節206R、206Lと、脛部207R、207Lと、足首208R、208Lと、足平209R、209Lとで構成され、股関節210R、210Lによって体幹部202の略最下端にて連結されている。また、左右各々の上肢203R、203Lは、上腕211R、211Lと、肘関節212R、212Lと、前腕213R、213Lとで構成され、肩関節21

4 R, 2 1 4 Lによって体幹部 2 0 2 上方の左右各側縁にて連結されている。また、頭部 2 0 4 は、首関節 2 5 5 によって体幹部 2 0 2 の略最上端中央に連結されている。

頭部 2 0 4 を支持する首関節ヨー軸 3 0 2 と、首関節ピッチ軸 3 0 3 と、首関節ロール軸 3 0 4 という 3 自由度を有している。

また、腕の関節は、肩関節ピッチ軸 3 0 8 と、肩関節ロール軸 3 0 9 と、上腕ヨー軸 3 1 0 と、肘関節ピッチ軸 3 1 1 と、前腕ヨー軸 3 1 2 と、手首関節ピッチ軸 3 1 3 と、手首関節ロール軸 3 1 4 と、手部 3 1 5 とで構成される。手部 3 1 5 は、実際には、複数本の指を含む多関節・多自由度構造体である。ただし、手部 3 1 5 の動作は人間型ロボット装置 2 0 0 の姿勢制御や歩行制御に対する寄与や影響が少ないので、本明細書ではゼロ自由度と仮定する。したがって、各腕部は 7 自由度を有するとする。

また、体幹部は、体幹ピッチ軸 3 0 5 と、体幹ロール軸 3 0 6 と、体幹ヨー軸 3 0 7 という 3 自由度を有する。

また、下肢を構成する各々の脚部は、股関節ヨー軸 3 1 6 と、股関節ピッチ軸 3 1 7 と、股関節ロール軸 3 1 8 と、膝関節ピッチ軸 3 1 9 と、足首関節ピッチ軸 3 2 0 と、足首関節ロール軸 3 2 1 と、足部 3 2 2 とで構成される。本明細書中では、股関節ピッチ軸 3 1 7 と股関節ロール軸 3 1 8 の交点は、人間型ロボット装置 2 0 0 の股関節位置を定義する。人体の足部 3 2 2 は、実際には多関節・多自由度の足底を含んだ構造体であるが、人間型ロボット装置 2 0 0 の足底は、ゼロ自由度とする。したがって、各脚部は、6 自由度で構成される。

以上を総括すれば、人間型ロボット装置 2 0 0 全体としては、合計で $3 + 7 \times 2 + 3 + 6 \times 2 = 32$ 自由度を有することになる。ただし、エンターテインメント向けの人間型ロボット装置 2 0 0 が必ずしも 32 自由度に限定される訳ではない。設計・制作上の制約条件や要求仕様等に応じて、自由度すなわち関節数を適宜増減することができるというまでもない。

上述したような人間型ロボット装置 2 0 0 がもつ各自由度は、実際にはアクチュエータを用いて実装される。外観上で余分な膨らみを排してヒトの自然体形状に近似させること、2 足歩行という不安定な構造体に対して姿勢制御を行うこと

などの要請から、アクチュエータは小型且つ軽量であることが好ましい。

図19には、人間型ロボット装置200の制御システム構成を模式的に示している。同図に示すように、人間型ロボット装置200は、ヒトの四肢を表現した各機構ユニット330、340、350R/L、360R/Lと、各機構ユニット間の協調動作を実現するための適応制御を行う制御ユニット380とで構成される（ただし、R及びLの各々は、右及び左の各々を示す接尾辞である。以下同様）。

人間型ロボット装置200全体の動作は、制御ユニット380によって統括的に制御される。制御ユニット380は、CPU (Central Processing Unit) やメモリ等の主要回路コンポーネント(図示しない)で構成される主制御部381と、電源回路や人間型ロボット装置200の各構成要素とのデータやコマンドの授受を行うインターフェイス（何れも図示しない）などを含んだ周辺回路382とで構成される。この制御ユニット380の設置場所は、特に限定されない。図19では体幹部ユニット340に搭載されているが、頭部ユニット330に搭載してもよい。或いは、人間型ロボット装置200外に制御ユニット380を配備して、人間型ロボット装置200の機体とは有線若しくは無線で交信するようにしてもよい。

図19に示した人間型ロボット装置200内の各関節自由度は、それぞれに対応するアクチュエータによって実現される。すなわち、頭部ユニット330には、首関節ヨー軸302、首関節ピッチ303、首関節ロール軸304の各々を表現する首関節ヨー軸アクチュエータ A_2 、首関節ピッチ軸アクチュエータ A_3 、首関節ロール軸アクチュエータ A_4 が配設されている。

また、頭部ユニット330には、外部の状況を撮像するためのCCD (Charge Coupled Device) カメラが設けられているほか、前方に位置する物体までの距離を測定するための距離センサ、外部音を集音するためのマイク、音声を出力するためのスピーカ、使用者からの「撫でる」や「叩く」といった物理的な働きかけにより受けた圧力を検出するためのタッチセンサ等が配設されている。

また、体幹部ユニット340には、体幹ピッチ軸305、体幹ロール軸306、体幹ヨー軸307の各々を表現する体幹ピッチ軸アクチュエータ A_5 、体幹ロー

ル軸アクチュエータ A_6 、体幹ヨー軸アクチュエータ A_7 が配設されている。また、体幹部ユニット 340 には、この人間型ロボット装置 200 の起動電源となるバッテリーを備えている。このバッテリーは、充放電可能な電池によって構成されている。

また、腕部ユニット 350 R/L は、上腕ユニット 351 R/L と、肘関節ユニット 352 R/L と、前腕ユニット 353 R/L に細分化されるが、肩関節ピッチ軸 308、肩関節ロール軸 309、上腕ヨー軸 310、肘関節ピッチ軸 311、前腕ヨー軸 312、手首関節ピッチ軸 313、手首関節ロール軸 314 の各々表現する肩関節ピッチ軸アクチュエータ A_8 、肩関節ロール軸アクチュエータ A_9 、上腕ヨー軸アクチュエータ A_{10} 、肘関節ピッチ軸アクチュエータ A_{11} 、肘関節ロール軸アクチュエータ A_{12} 、手首関節ピッチ軸アクチュエータ A_{13} 、手首関節ロール軸アクチュエータ A_{14} が配備されている。

また、脚部ユニット 360 R/L は、大腿部ユニット 361 R/L と、膝ユニット 362 R/L と、脛部ユニット 363 R/L に細分化されるが、股関節ヨー軸 316、股関節ピッチ軸 317、股関節ロール軸 318、膝関節ピッチ軸 319、足首関節ピッチ軸 320、足首関節ロール軸 321 の各々を表現する股関節ヨー軸アクチュエータ A_{16} 、股関節ピッチ軸アクチュエータ A_{17} 、股関節ロール軸アクチュエータ A_{18} 、膝関節ピッチ軸アクチュエータ A_{19} 、足首関節ピッチ軸アクチュエータ A_{20} 、足首関節ロール軸アクチュエータ A_{21} が配備されている。各関節に用いられるアクチュエータ A_2 、 $A_3 \cdots$ は、より好ましくは、ギア直結型で且つサーボ制御系をワンチップ化してモータ・ユニット内に搭載したタイプの小型 AC サーボ・アクチュエータで構成することができる。

頭部ユニット 330、体幹部ユニット 340、腕部ユニット 350、各脚部ユニット 360 などの各機構ユニット毎に、アクチュエータ駆動制御部の副制御部 335、345、355 R/L、365 R/L が配備されている。更に、各脚部 360 R、L の足底が着床したか否かを検出する接地確認センサ 391 及び 392 を装着するとともに、体幹部ユニット 340 内には、姿勢を計測する姿勢センサ 393 を装備している。

接地確認センサ 391 及び 392 は、例えば足底に設置された近接センサ又は

マイクロ・スイッチなどで構成される。また、姿勢センサ 393 は、例えば、加速度センサとジャイロ・センサの組み合わせによって構成される。

接地確認センサ 391 及び 392 の出力によって、歩行・走行などの動作期間中において、左右の各脚部が現在立脚又は遊脚何れの状態であるかを判別することができる。また、姿勢センサ 393 の出力により、体幹部分の傾きや姿勢を検出することができる。

主制御部 381 は、各センサ 391 ~ 393 の出力に応答して制御目標をダイナミックに補正することができる。より具体的には、副制御部 335, 345, 355 R/L, 365 R/L の各々に対して適応的な制御を行い、人間型ロボット装置 200 の上肢、体幹、及び下肢が協調して駆動する全身運動パターンを実現できる。

人間型ロボット装置 200 の機体上での全身運動は、足部運動、ZMP (Zero Moment Point) 軌道、体幹運動、上肢運動、腰部高さなどを設定するとともに、これらの設定内容にしたがった動作を指示するコマンドを各副制御部 335, 345, 355 R/L, 365 R/L に転送する。そして、各々の副制御部 335, 345, ... 等では、主制御部 381 からの受信コマンドを解釈して、各アクチュエータ A_2 , A_3 , ... 等に対して駆動制御信号を出力する。ここでいう「ZMP」とは、歩行中の床反力によるモーメントがゼロとなる床面上の点のことであり、また、「ZMP 軌道」とは、例えば人間型ロボット装置 200 の歩行動作期間中に ZMP が動く軌跡を意味する。

歩行時には、重力と歩行運動に伴って生じる加速度によって、歩行系から路面には重力と慣性力、並びにこれらのモーメントが作用する。いわゆる「ダランベールの原理」によると、それらは路面から歩行系への反作用としての床反力、床反力モーメントとバランスする。力学的推論の帰結として、足底接地点と路面の接触形成する支持多角形の辺上或いはその内側にピッチ及びロール軸モーメントがゼロとなる点、すなわち「ZMP (Zero Moment Point)」が存在する。

脚式移動ロボットの姿勢安定制御や歩行時の転倒防止に関する提案の多くは、この ZMP を歩行の安定度判別の規範として用いたものである。ZMP 規範に基づく 2 足歩行パターン生成は、足底着地点を予め設定することができ、路面形状

に応じた足先の運動学的拘束条件を考慮しやすいなどの利点がある。また、ZMPを安定度判別規範とすることは、力ではなく軌道を運動制御上の目標値として扱うことを意味するので、技術的に実現可能性が高まる。なお、ZMPの概念並びにZMPを歩行ロボットの安定度判別規範に適用する点については、Miomir Vukobratovic 著“LEGGED LOCOMOTION ROBOTS”（加藤一郎外著『歩行ロボットと人工の足』（日刊工業新聞社））に記載されている。

一般には、4足歩行よりもヒューマノイドのような2足歩行のロボットの方が、重心位置が高く、且つ、歩行時のZMP安定領域が狭い。したがって、このような路面状態の変化に伴う姿勢変動の問題は、2足歩行ロボットにおいてとりわけ重要となる。

以上のように、人間型ロボット装置200は、各々の副制御部335, 345, ...等が、主制御部381からの受信コマンドを解釈して、各アクチュエータ A_2 , A_3 , ...に対して駆動制御信号を出力し、各ユニットの駆動を制御している。これにより、人間型ロボット装置200は、目標の姿勢に安定して遷移し、安定した姿勢で歩行できる。

また、人間型ロボット装置200における制御ユニット380では、上述したような姿勢制御のほかに、加速度センサ、タッチセンサ、接地確認センサ等の各種センサ、及びCCDカメラからの画像情報、マイクからの音声情報等を統括して処理している。制御ユニット380では、図示しないが加速度センサ、ジャイロ・センサ、タッチセンサ、距離センサ、マイク、スピーカなどの各種センサ、各アクチュエータ、CCDカメラ及びバッテリーが各々対応するハブを介して主制御部381と接続されている。

主制御部381は、上述の各センサから供給されるセンサデータや画像データ及び音声データを順次取り込み、これらをそれぞれ内部インターフェイスを介してDRAM内の所定位置に順次格納する。また、主制御部381は、バッテリーから供給されるバッテリー残量を表すバッテリー残量データを順次取り込み、これをDRAM内の所定位置に格納する。DRAMに格納された各センサデータ、画像データ、音声データ及びバッテリー残量データは、主制御部381がこの人間型ロボット装置200の動作制御を行う際に利用される。

主制御部 381 は、人間型ロボット装置 200 の電源が投入された初期時、制御プログラムを読み出し、これを DRAM に格納する。また、主制御部 381 は、上述のように主制御部 381 より DRAM に順次格納される各センサデータ、画像データ、音声データ及びバッテリー残量データに基づいて自己及び周囲の状況や、使用者からの指示及び働きかけの有無などを判断する。更に、主制御部 381 は、この判断結果及び DRAM に格納した制御プログラムに基づいて自己の状況に応じて行動を決定するとともに、当該決定結果に基づいて必要なアクチュエータを駆動させることにより人間型ロボット装置 200 に、いわゆる「身振り」、「手振り」といった行動をとらせる。

したがって、人間型ロボット装置 200 は、制御プログラムに基づいて自己及び周囲の状況を判断し、使用者からの指示及び働きかけに応じて自律的に行動できる。また、人間型ロボット装置 200 は、CCD カメラにおいて撮像された画像から抽出した文字の発音のしかた（読み方）を、抽出された文字から推定される読み方と集音マイクにおいて集音された音声とをマッチングして決定する。したがって、人間型ロボット装置 200 の音声認識の精度が向上し、新規単語が音声認識用辞書に登録できる。

産業上の利用可能性

以上詳細に説明したように、本発明にかかるロボット装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から文字を抽出する文字認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、文字認識手段において抽出された文字から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声

波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

本発明にかかるロボット装置は、撮像手段において撮像された画像から抽出された文字から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与した複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成し、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較して最も近い音声波形を抽出した文字の発音のしかたであるとして決定する。

したがって、本発明にかかるロボット装置によれば、特に、弱い音素（例えば、語頭の／s／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上できる。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

また、本発明にかかるロボット装置は、単語とこの単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、記憶制御手段が新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて音声認識用辞書に記憶する。

したがって、本発明にかかるロボット装置によれば、入力した音声及び出力する音声に文法規則、対話規則等を適用する上で必要となる単語属性情報をユーザが入力する必要がなくなり利便性が向上するとともに、ユーザが属性情報を知らない場合は、属性情報を入力できなかったという不都合が改善される。

また、本発明にかかる文字認識装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から文字を抽出する文字認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、文字認識手段において抽出された文字から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声

波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

したがって、本発明にかかる文字認識装置によれば、特に、弱い音素（例えば、語頭の／s／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上できる。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

また、本発明にかかる文字認識装置は、単語とこの単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、記憶制御手段が新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて音声認識用辞書に記憶する。

したがって、本発明にかかる文字認識装置によれば、入力した音声及び出力する音声に文法規則、対話規則等を適用する上で必要となる単語属性情報をユーザが入力する必要がなくなり利便性が向上するとともに、ユーザが属性情報を知らない場合は、属性情報を入力できなかったという不都合が改善される。

また、本発明にかかる文字認識方法は、被写体を撮像する撮像工程と、撮像工程において撮像された画像から文字を抽出する文字認識工程と、周囲の音を取得する集音工程と、集音工程において取得された音から音声を認識する音声認識工程と、文字認識工程において抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成工程と、発音情報生成工程において生成された各音声波形と音声認識工程において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶制御工程とを備える。

したがって、本発明にかかる文字認識方法によれば、特に、弱い音素（例えば、

語頭の／s／等)を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上できる。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

また、本発明にかかる文字認識方法によれば、単語とこの単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、記憶制御手段が新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて音声認識用辞書に記憶する。

したがって、本発明にかかる文字認識方法によれば、入力した音声及び出力する音声に文法規則、対話規則等を適用する上で必要となる単語属性情報をユーザが入力する必要がなくなり利便性が向上するとともに、ユーザが属性情報を知らない場合は、属性情報を入力できなかったという不都合が改善される。

更に、本発明にかかる制御プログラムは、被写体を撮像する撮像処理と、撮像処理によって撮像された画像から文字を抽出する文字認識処理と、周囲の音を取得する集音処理と、集音処理によって取得された音から音声を認識する音声認識処理と、文字認識処理によって抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、発音情報生成処理によって生成された各音声波形と音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理とをロボット装置に実行させる。

したがって、本発明にかかる制御プログラムによれば、ロボット装置は、特に、弱い音素（例えば、語頭の／s／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上される。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

また、上述の制御プログラムを記録媒体に記録して提供することによって、この記録媒体を読込可能で画像認識手段と音声認識手段とを備える音声認識装置としての機能を有する電子機器に対して、新規単語を登録する際の認識精度が向上される。これにより、正確な発音のしかたが記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

請求の範囲

1. 内部状態に応じて自律的に動作するロボット装置において、

単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、

単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、

被写体を撮像する撮像手段と、

上記撮像手段において撮像された画像から文字を抽出する文字認識手段と、
周囲の音を取得する集音手段と、

上記集音手段において取得された音から音声を認識する音声認識手段と、

上記文字認識手段において抽出された文字から推定される複数通りの表音文字を上記単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、

上記発音情報生成手段において生成された各音声波形と上記音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして上記音声認識用辞書に新規に記憶する記憶制御手段と

を備えることを特徴とするロボット装置。

2. 上記画像から抽出される複数個の文字と該文字に対して付与される複数通りの発音のしかたとの対応を一時辞書として一時的に記憶する一時記憶手段を備えることを特徴とする請求の範囲第1項記載のロボット装置。

3. 単語と該単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、記憶制御手段は、新規に記憶する文字と該文字の発音のしかたとともに上記単語属性を対応させて上記音声認識用辞書に記憶することを特徴とする請求の範囲第1項記載のロボット装置。

4. 上記音声認識手段において認識された音声に対する応答を生成する対話管理

手段を備え、上記対話管理手段は、上記単語属性を音声に対する応答規則で使用することを特徴とする請求の範囲第3項記載のロボット装置。

5. 上記音声認識手段は、隠れマルコフモデル法に基づいて音声を認識することを特徴とする請求の範囲第1項記載のロボット装置。

6. 単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、

単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、

被写体を撮像する撮像手段と、

上記撮像手段において撮像された画像から文字を抽出する文字認識手段と、

周囲の音を取得する集音手段と、

上記集音手段において取得された音から音声を認識する音声認識手段と、

上記文字認識手段において抽出された文字から推定される複数通りの表音文字を上記単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、

上記発音情報生成手段において生成された各音声波形と上記音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして上記音声認識用辞書に新規に記憶する記憶制御手段と

を備えることを特徴とする文字認識装置。

7. 上記画像から抽出される複数個の文字と該文字に対して付与される複数通りの発音のしかたとの対応を一時辞書として一時的に記憶する一時記憶手段を備えることを特徴とする請求の範囲第6項記載の文字認識装置。

8. 単語と該単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、記憶制御手段は、新規に記憶する文字と該文字の発音のしかたとともに上記単語属性を対応させて上記音声認識用辞書に記憶することを特徴とする請求の範囲第6項記載の文字認識装置。

9. 上記音声認識手段において認識された音声に対する応答を生成する対話管理

手段を備え、上記対話管理手段は、上記単語属性を音声に対する応答規則で使用することを特徴とする請求の範囲第 8 項記載の文字認識装置。

10. 上記音声認識手段は、隠れマルコフモデル法に基づいて音声を認識することを特徴とする請求の範囲第 6 項記載の文字認識装置。

11. 被写体を撮像する撮像工程と、

上記撮像工程において撮像された画像から文字を抽出する文字認識工程と、
周囲の音を取得する集音工程と、

上記集音工程において取得された音から音声を認識する音声認識工程と、

上記文字認識工程において抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成工程と、

上記発音情報生成工程において生成された各音声波形と上記音声認識工程において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶制御工程と

を備えることを特徴とする文字認識方法。

12. 上記画像から抽出される複数個の文字と該文字に対して付与される複数通りの発音のしかたとの対応を一時辞書として一時記憶手段に記憶する工程を備えることを特徴とする請求の範囲第 11 項記載の文字認識方法。

13. 記憶制御工程では、新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて上記音声認識用辞書に記憶することを特徴とする請求の範囲第 11 項記載の文字認識方法。

14. 上記音声認識工程において認識された音声に対する応答を生成する対話管理工程を備え、上記対話管理工程では、上記単語属性が音声に対する応答規則で使用されることを特徴とする請求の範囲第 13 項記載の文字認識方法。

15. 上記音声認識工程では、隠れマルコフモデル法に基づいて音声認識されることを特徴とする請求の範囲第 11 項記載の文字認識方法。

16. 内部状態に応じて自律的に動作するロボット装置の制御プログラムにおい

て、

被写体を撮像する撮像処理と、

上記撮像処理によって撮像された画像から文字を抽出する文字認識処理と、

周囲の音を取得する集音処理と、

上記集音処理によって取得された音から音声を認識する音声認識処理と、

上記文字認識処理によって抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、

上記発音情報生成処理によって生成された各音声波形と上記音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理と

をロボット装置に実行させることを特徴とする制御プログラム。

17. 被写体を撮像する撮像処理と、

上記撮像処理によって撮像された画像から文字を抽出する文字認識処理と、

周囲の音を取得する集音処理と、

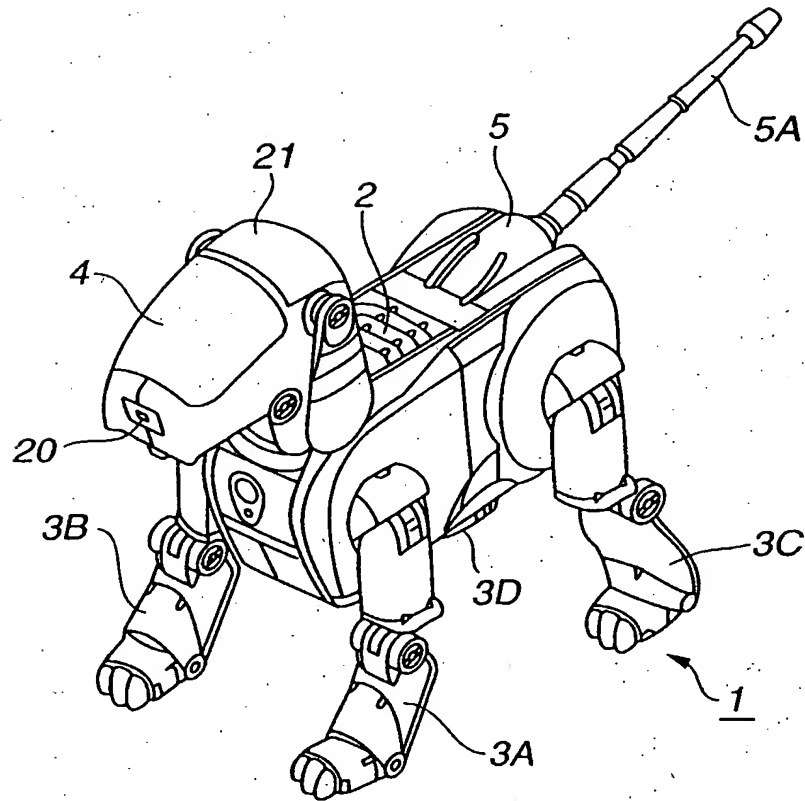
上記集音処理によって取得された音から音声を認識する音声認識処理と、

上記文字認識処理によって抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、

上記発音情報生成処理によって生成された各音声波形と上記音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理と

をロボット装置に実行させるための制御プログラムが記録された記録媒体。

1/19

**FIG.1**

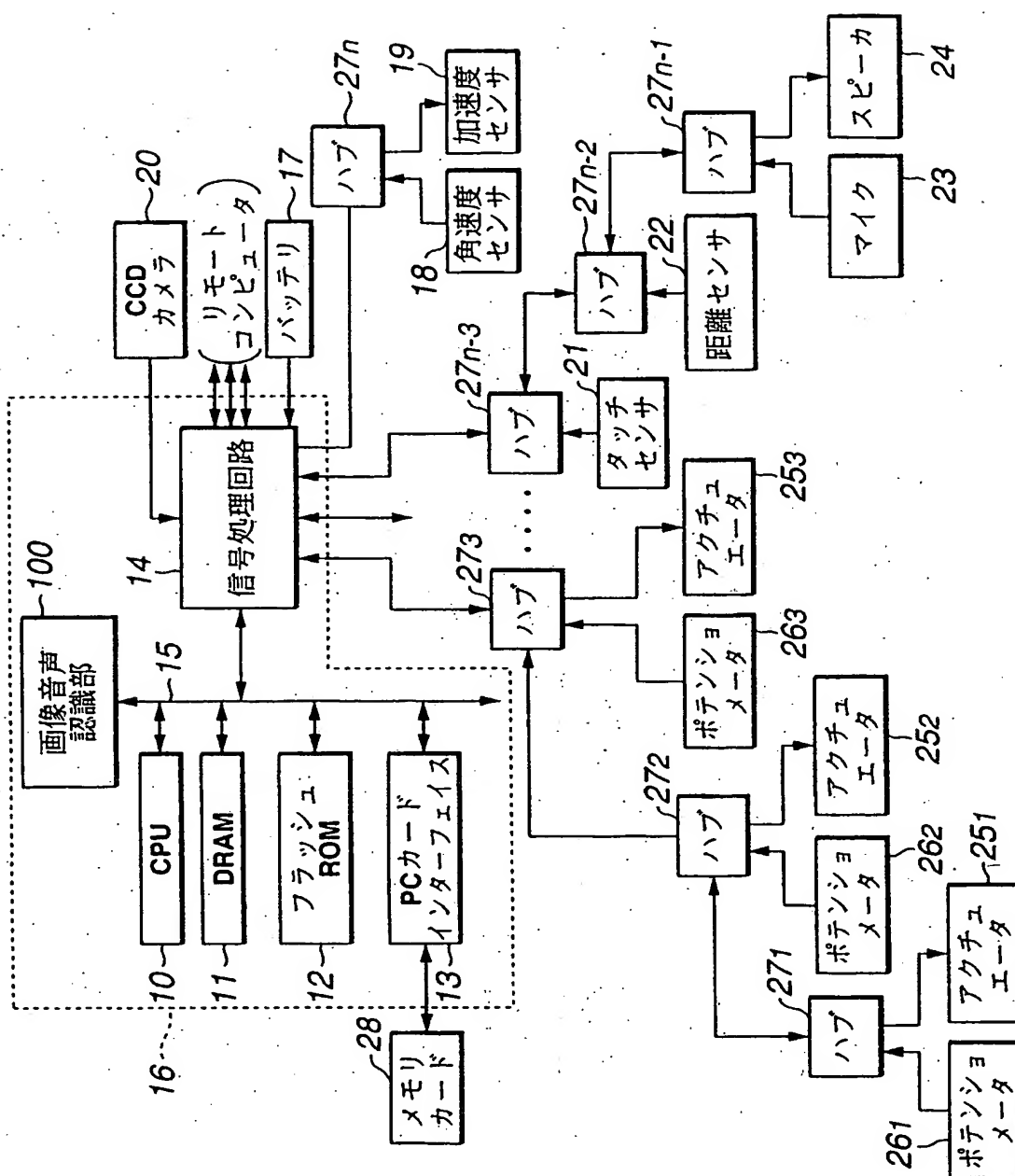


FIG. 2

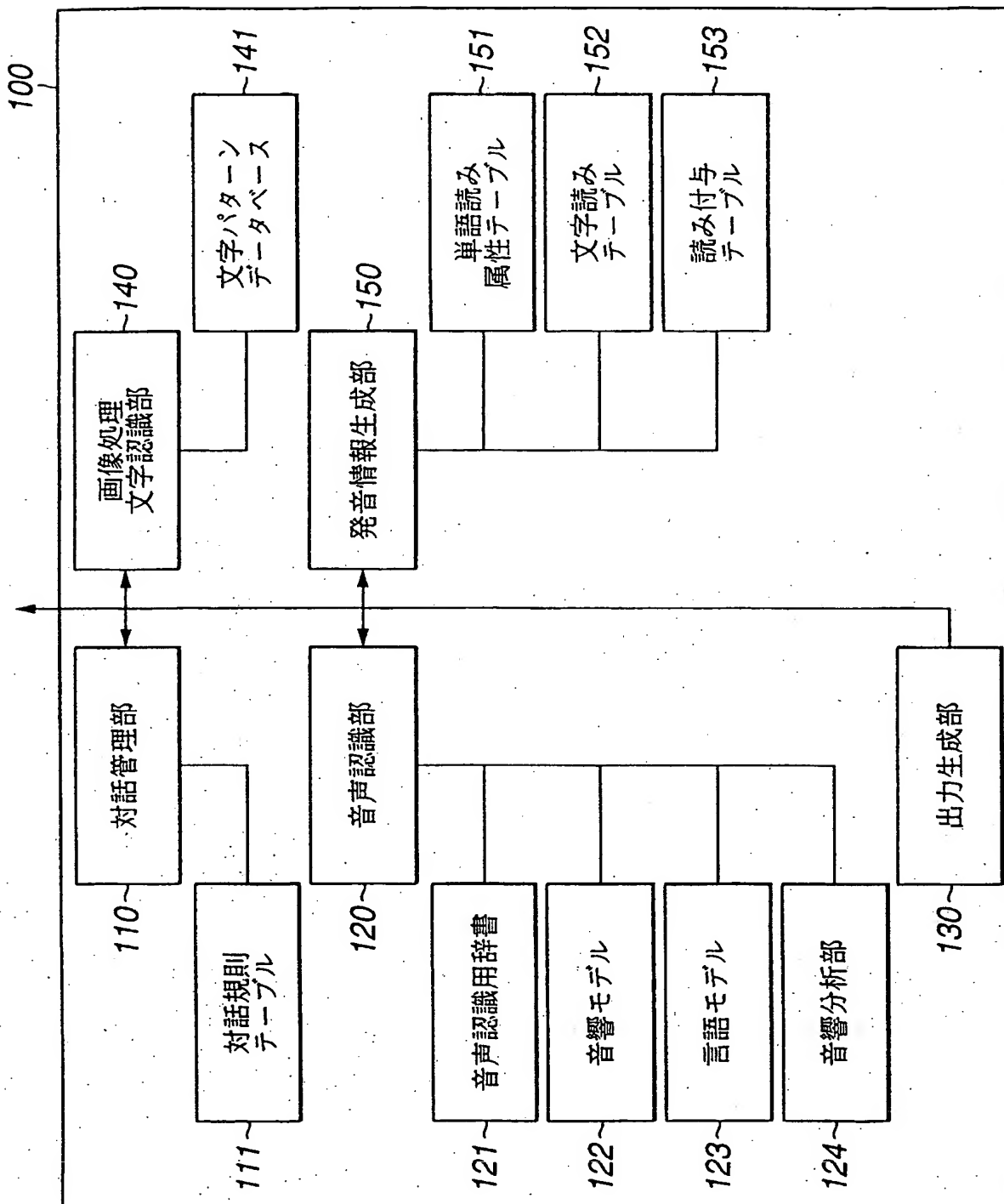


FIG.3

4/19

単語シンボル	PLU列
五反田	g o t a N d a
電車	d e N s h a
に	n i
行き	i k i
たい	t a i
⋮	⋮
⋮	⋮

FIG.4

単語	読み	属性
品川	しながわ	<地名>
北品川	きたしながわ	<地名>
青物横丁	あおものよこちょう	<地名>
⋮	⋮	⋮
佐藤	さとう	<姓>
鈴木	すずき	<姓>
⋮	⋮	⋮
すべすべ饅頭蟹	すべすべまんじゅうがに	<動物>
大錨海鼠	おおいかりなまこ	<動物>
⋮	⋮	⋮
⋮	⋮	⋮

FIG.5

5/19

単語	読み
北	きた、ホク
品	しな、ヒン、ホン
川	かわ、セン
比	くら、ヒ
晶	あきら、ショウ
・	・
・	・
・	・
?	はてな、クエスチョン
@	アット
・	・
・	・
・	・
A	エー、ア
B	ビー、ブ
・	・
・	・
・	・

FIG.6

6/19

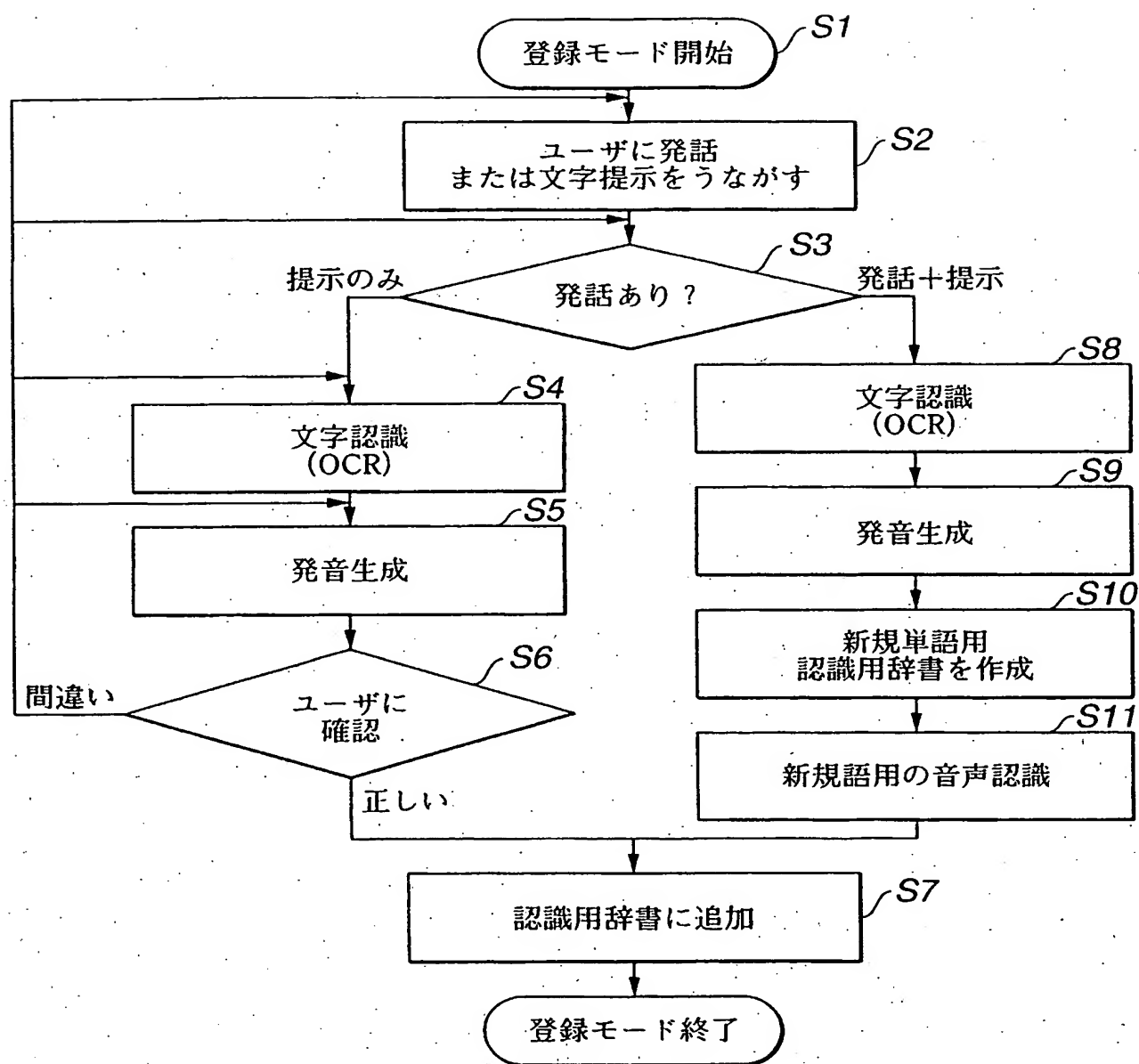


FIG.7

7/19

単語シンボル	PLU列
北品川	k i t a s i n a g a w a
北晶川	h i s h o : k a w a
比晶川	k u r a a k i r a g a w a

FIG.8

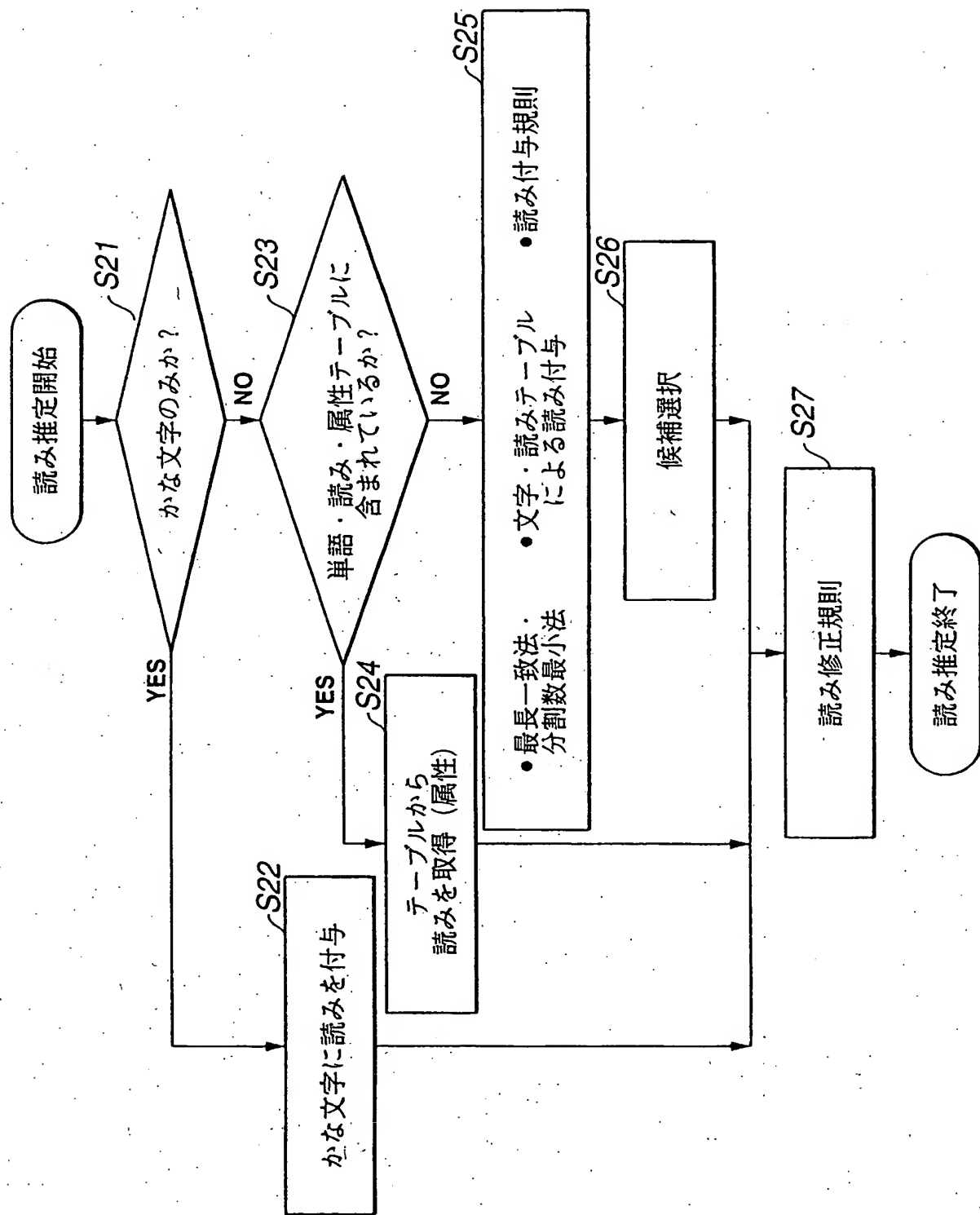


FIG. 9

9/19

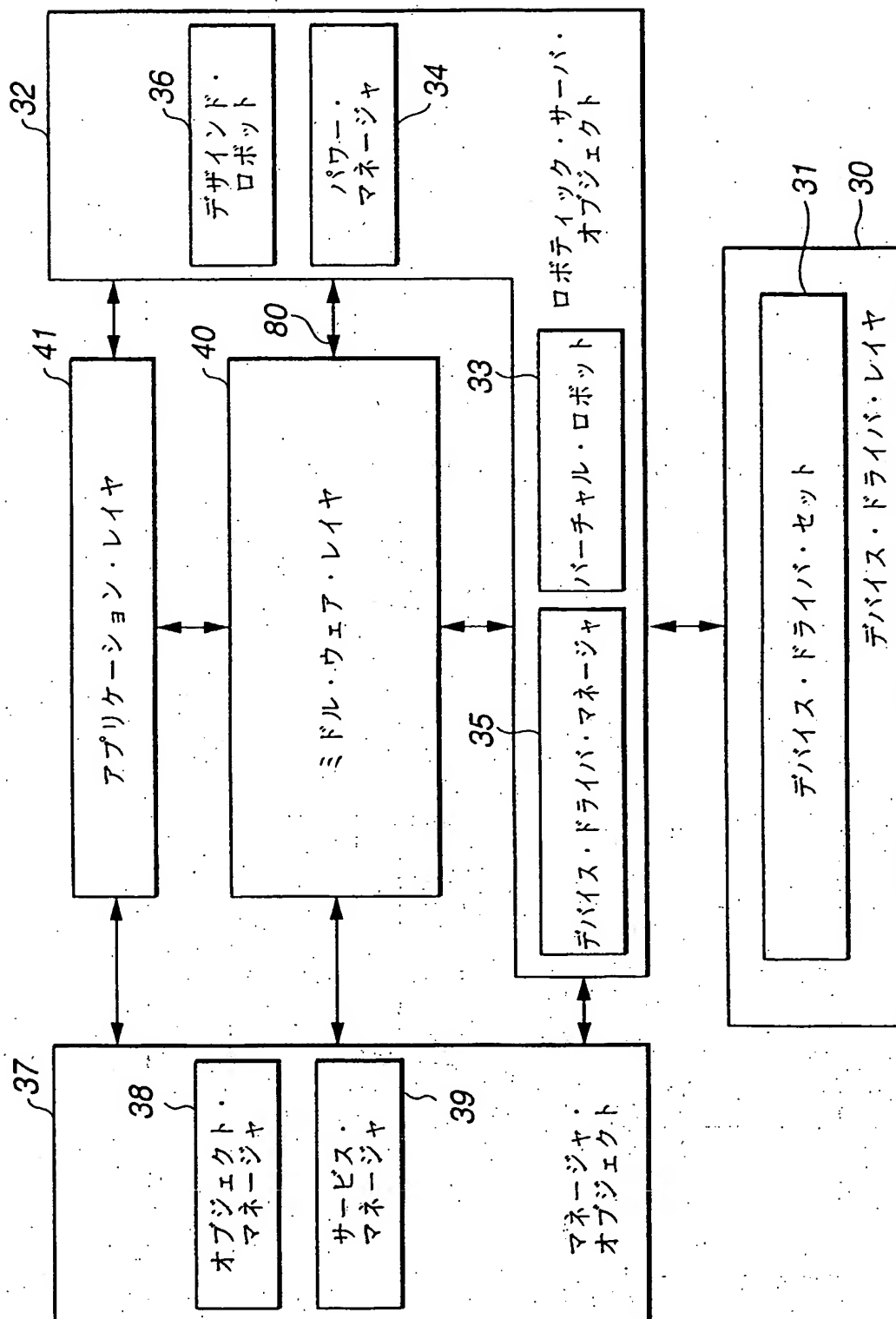


FIG.10

10/19

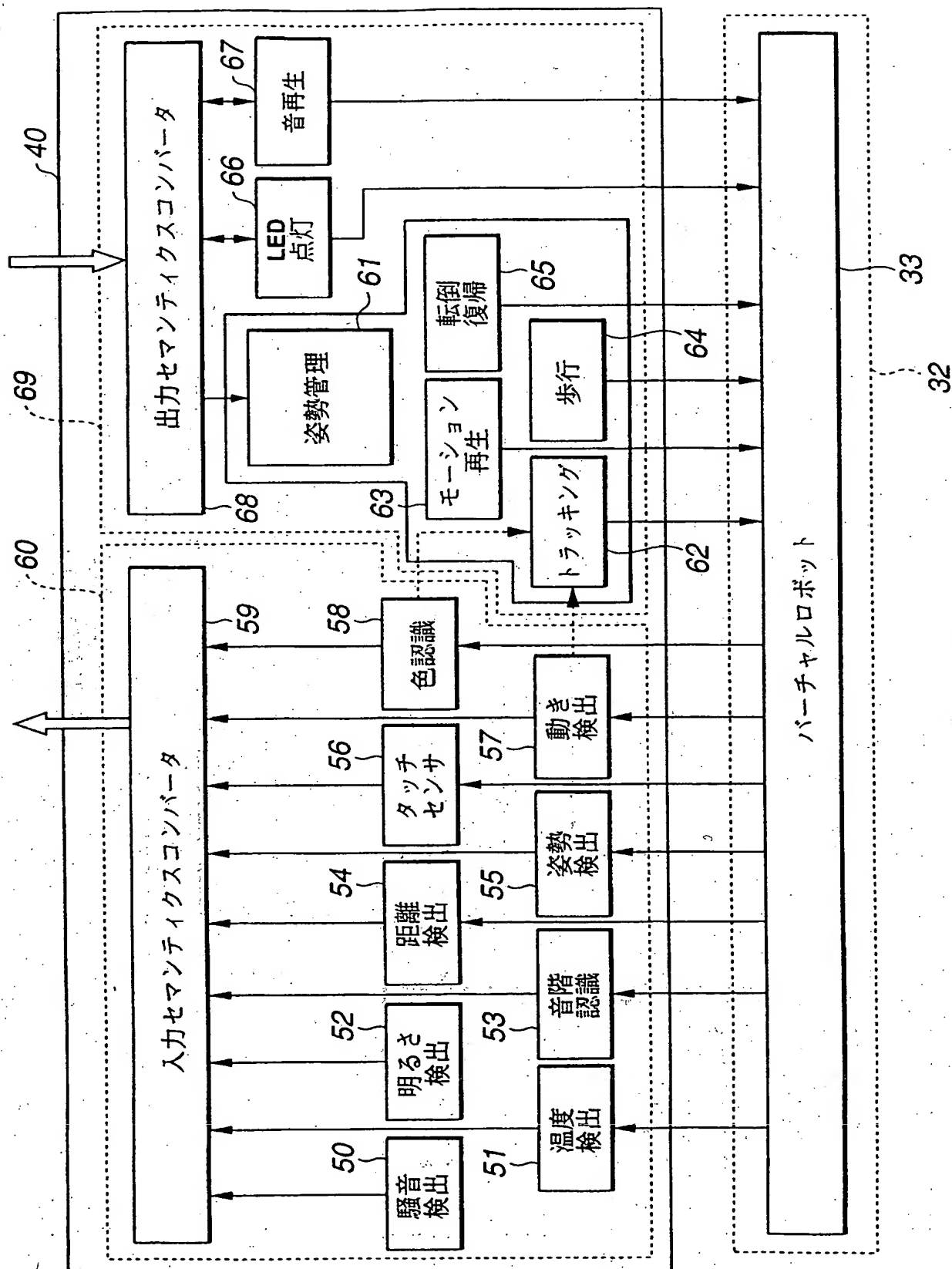


FIG.11

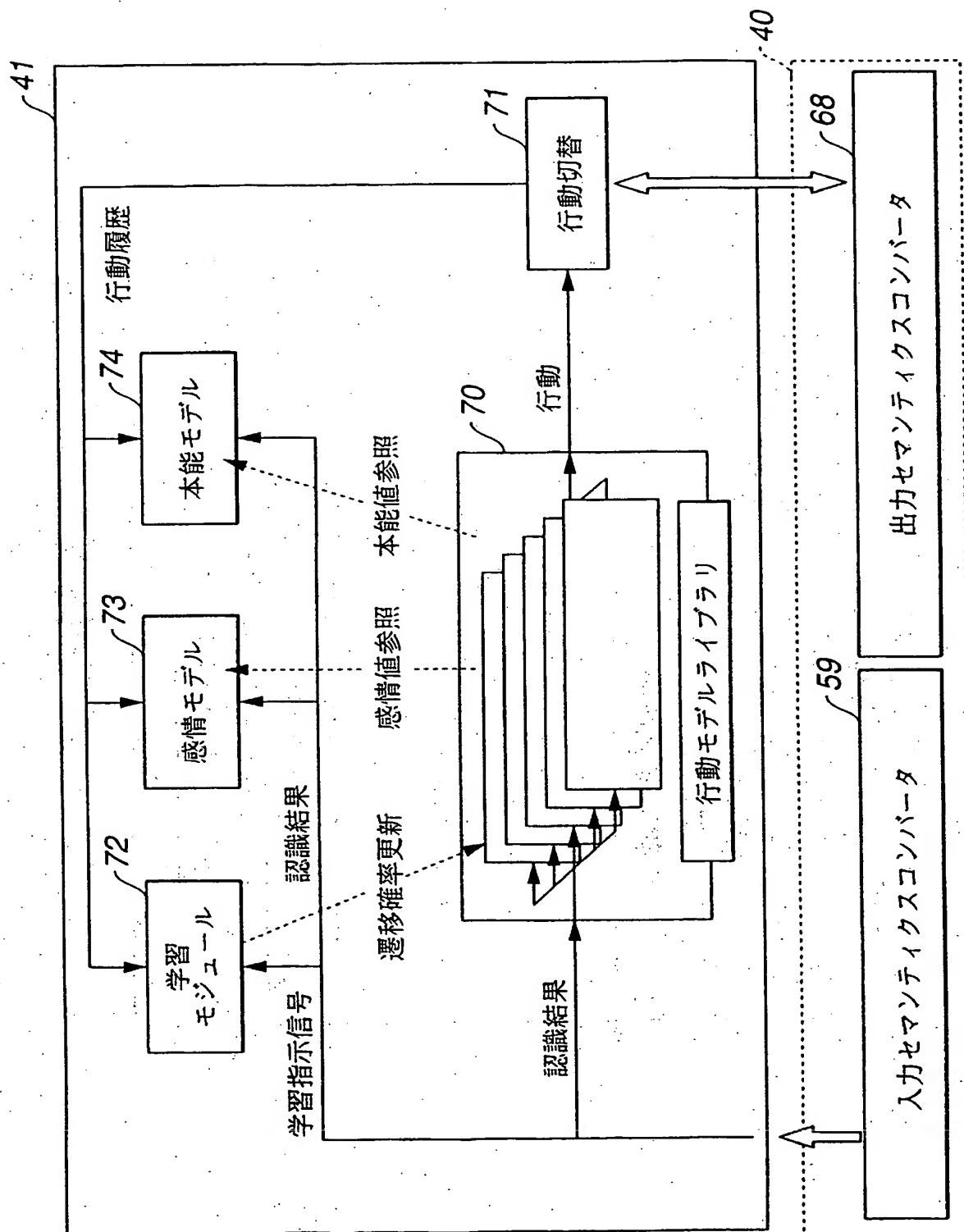


FIG.12

12/19

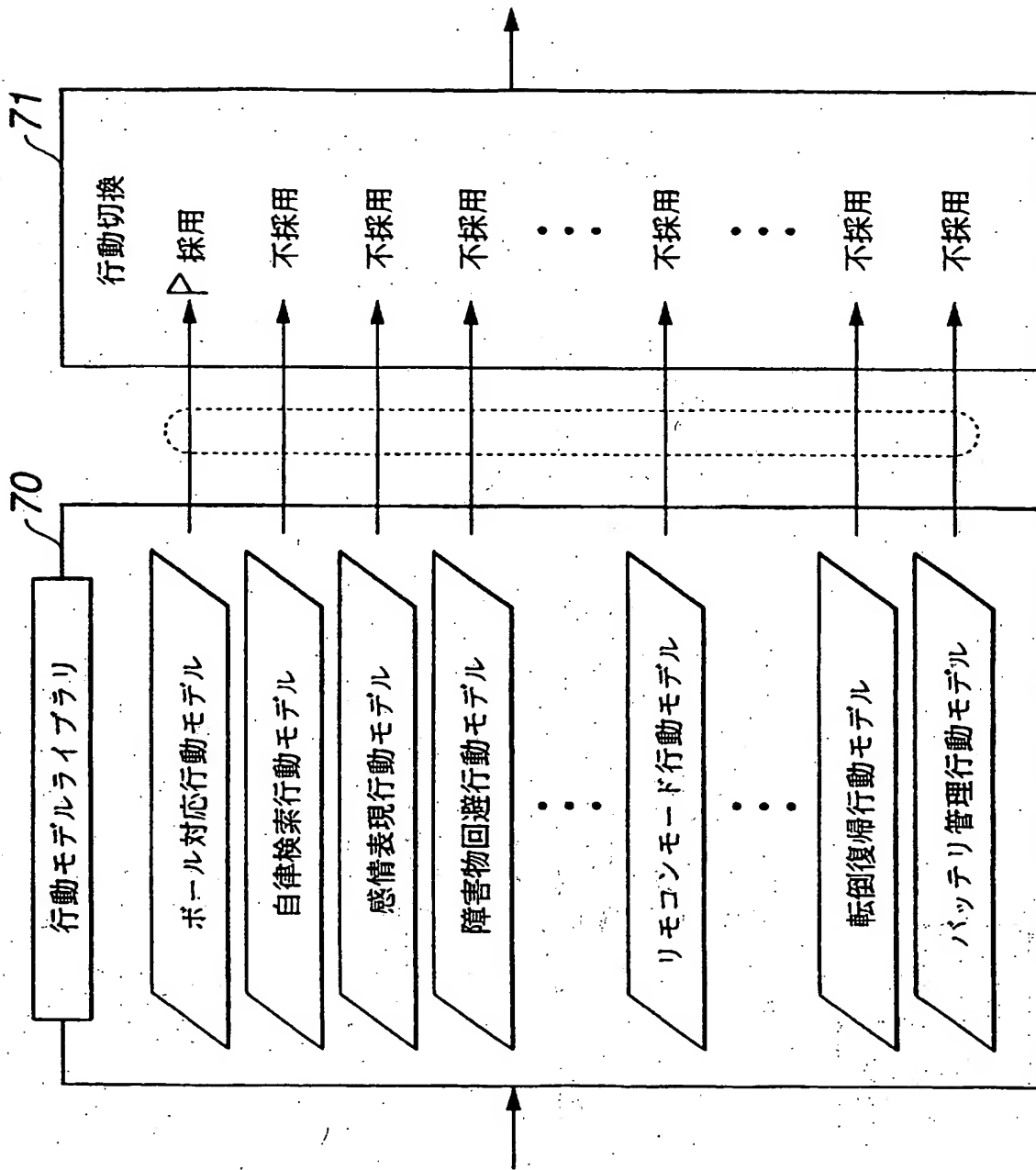
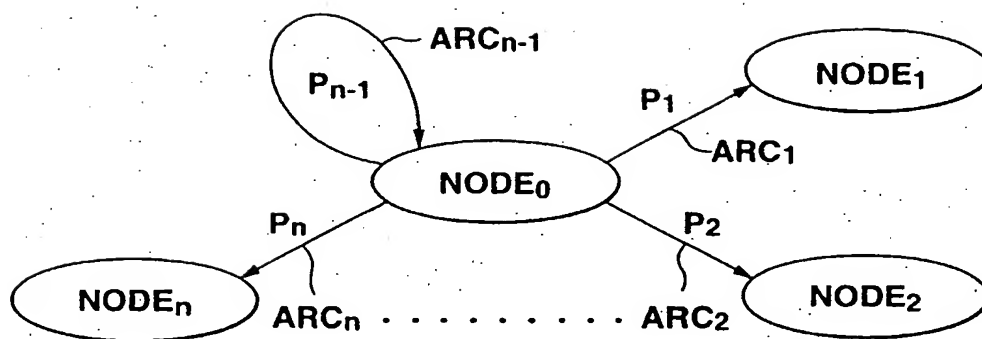


FIG.13

13/19

**FIG.14**

	入カイベント名	データ名	データの範囲	他のノードへの遷移確率 Di					
				A	B	C	D	n	
node 100								n	
遷移先ノード				node 120	node 120	node 1000		node 600	
出力行動				ACTION 1	ACTION 2	MOVE BACK		ACTION 4	
1	BALL	SIZE	0.1000	30%					
2	PAT				40%				
3	HIT				20%				
4	MOTION					50%			
5	OBSTACLE	DISTANCE	0.100			100%			
6		JOY	50.100						
7		SURPRISE	50.100						
8		SADNESS	50.100						

FIG.15

15/19

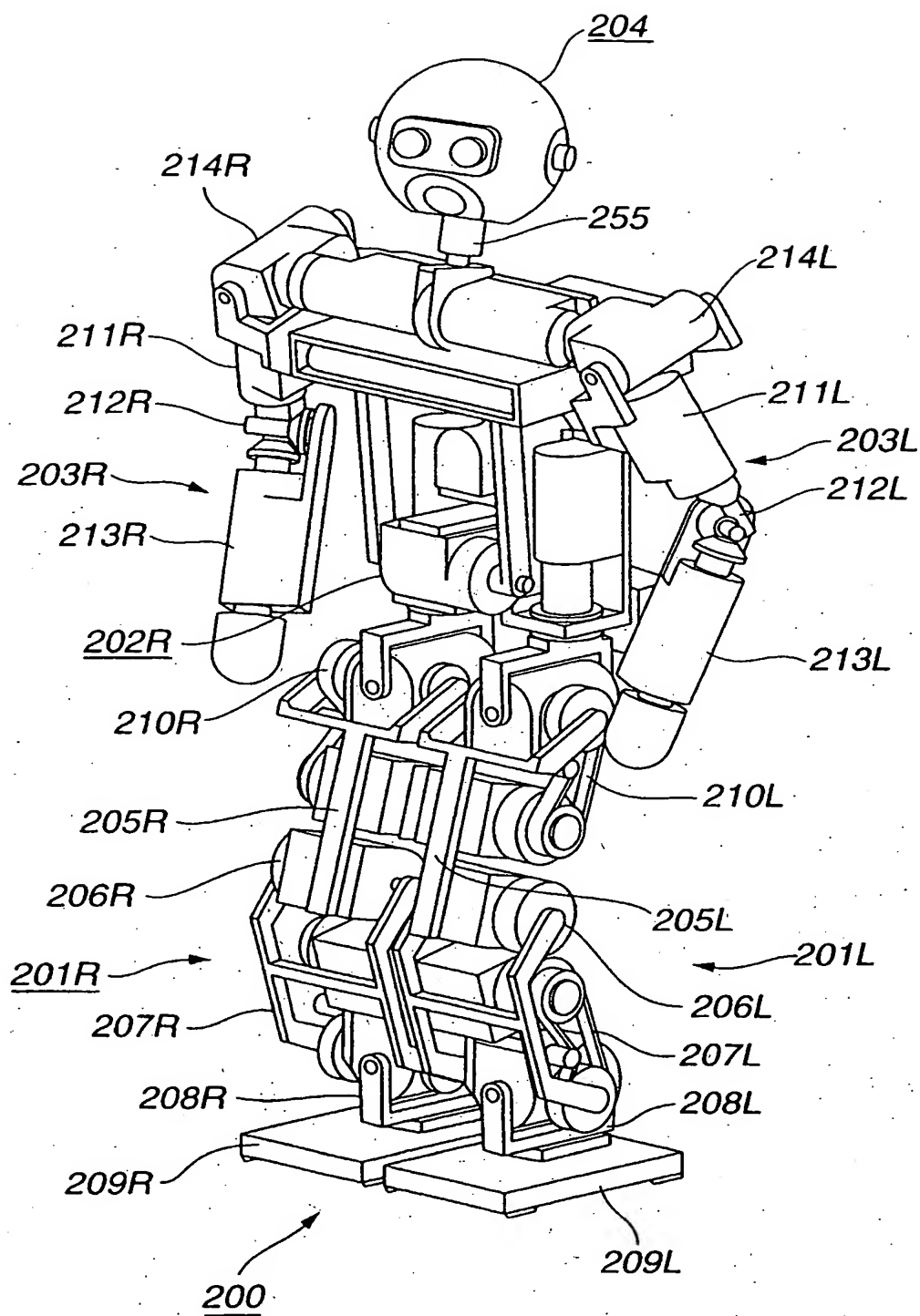


FIG. 16

16/19

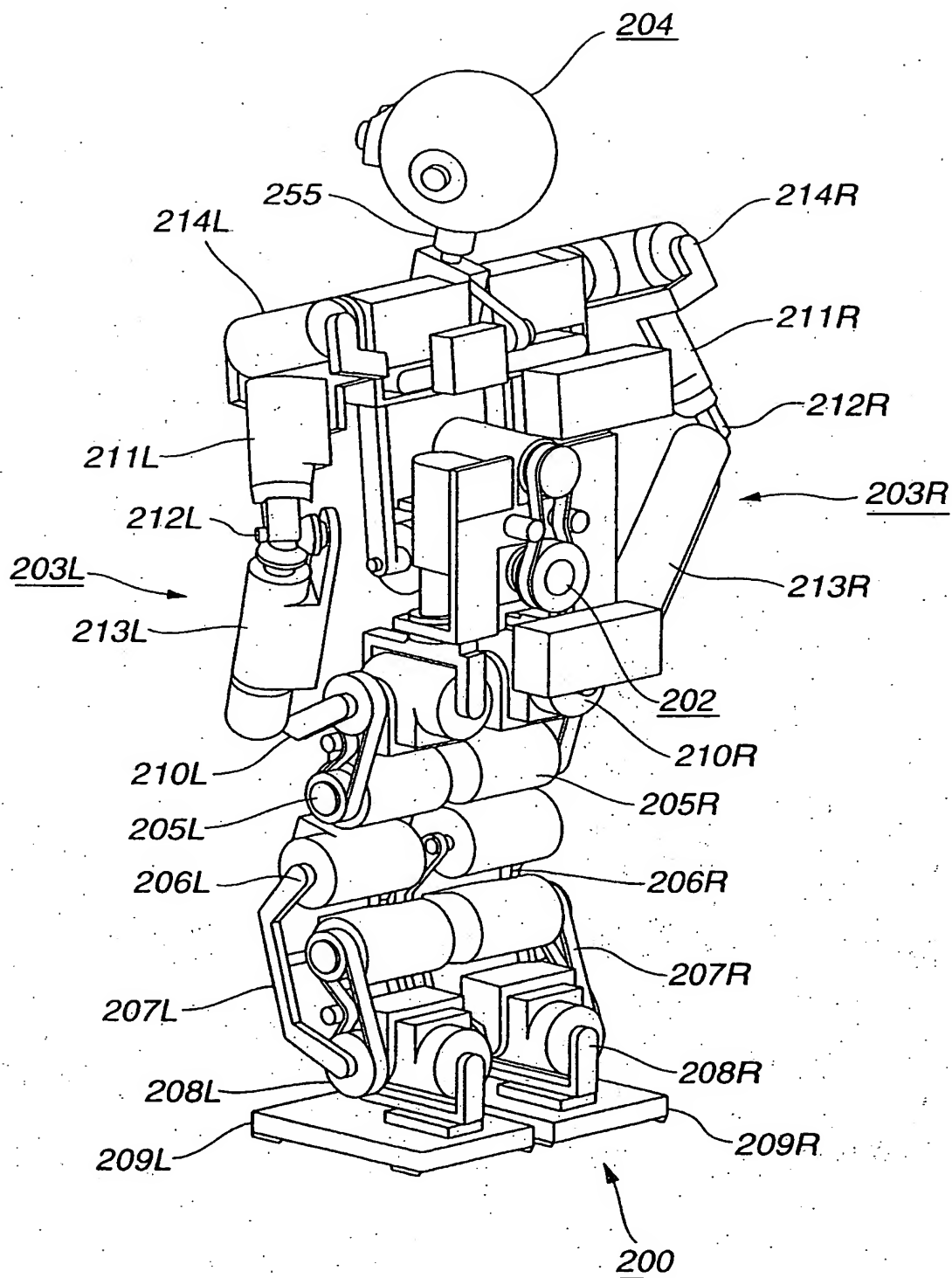
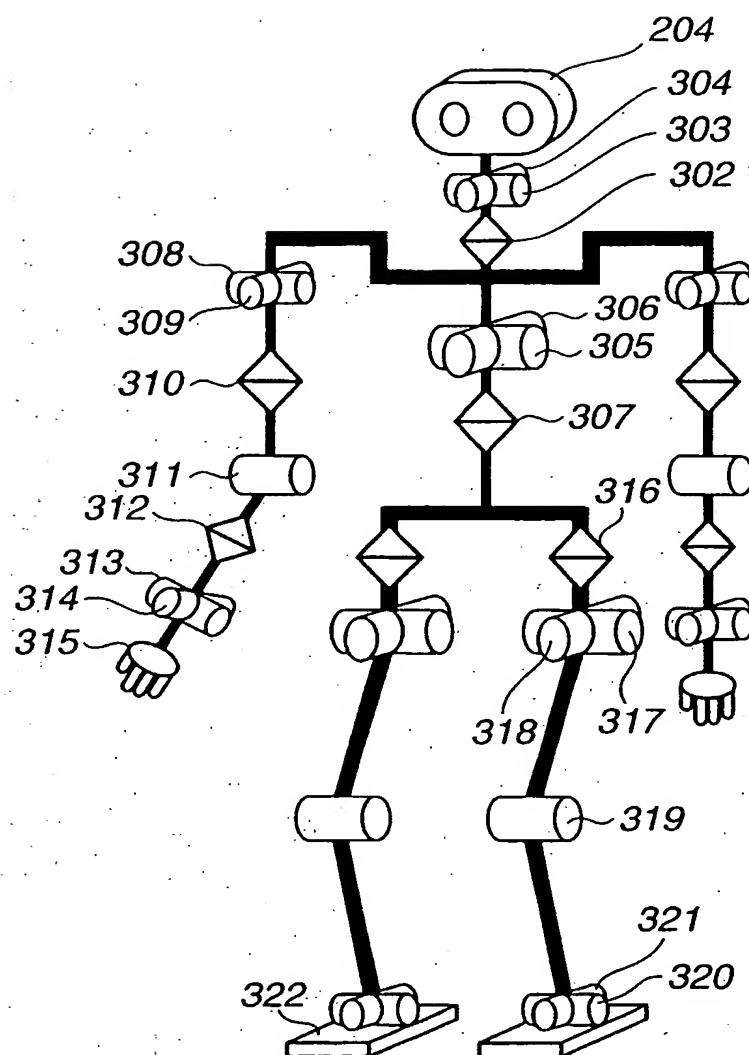


FIG.17

17/19

**FIG.18**

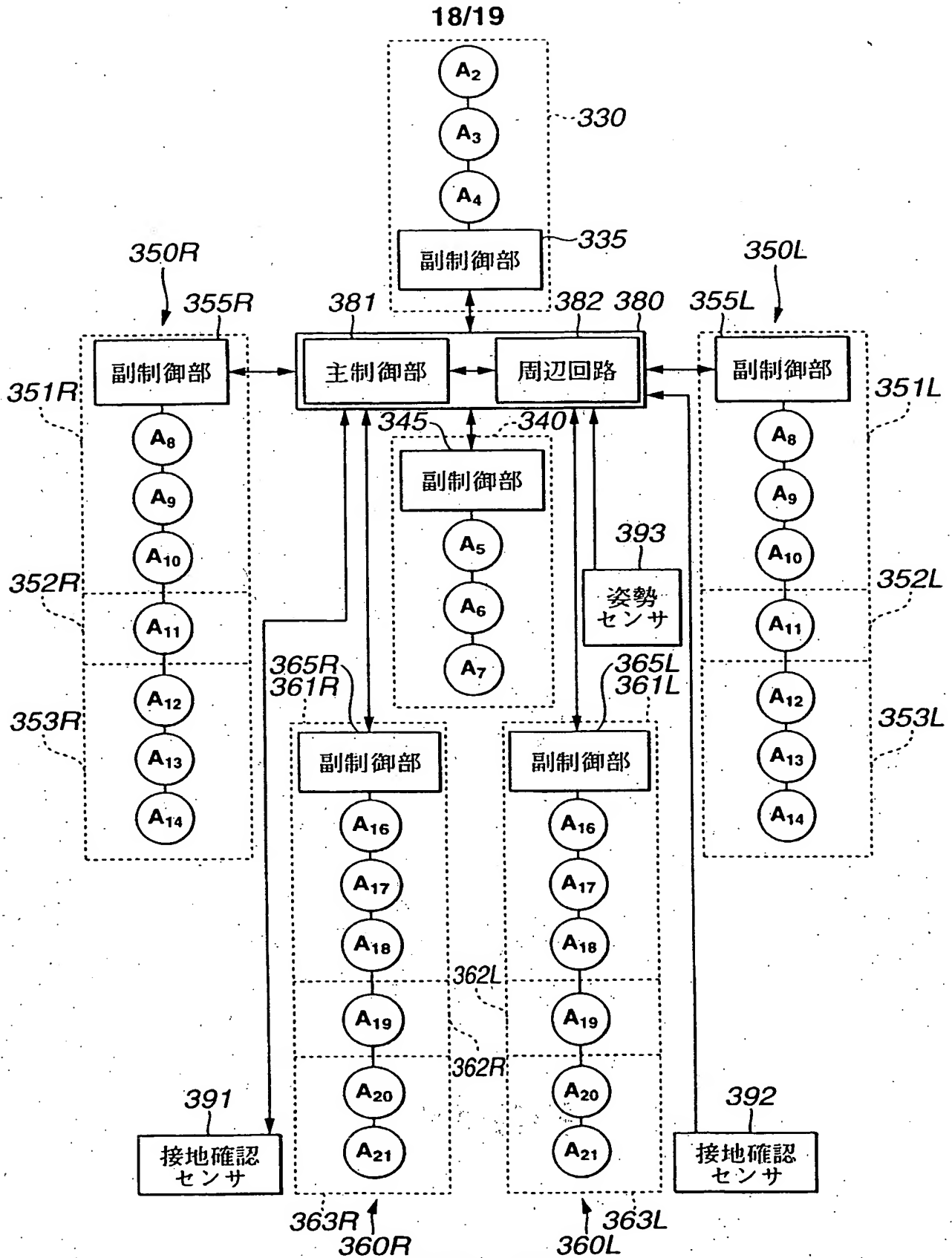


FIG. 19

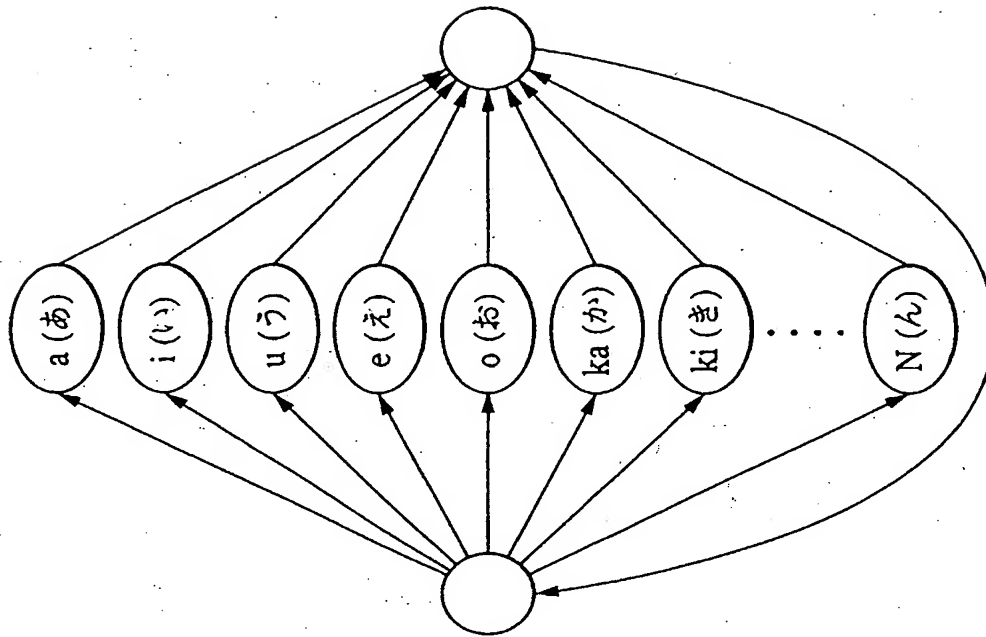


FIG.20A

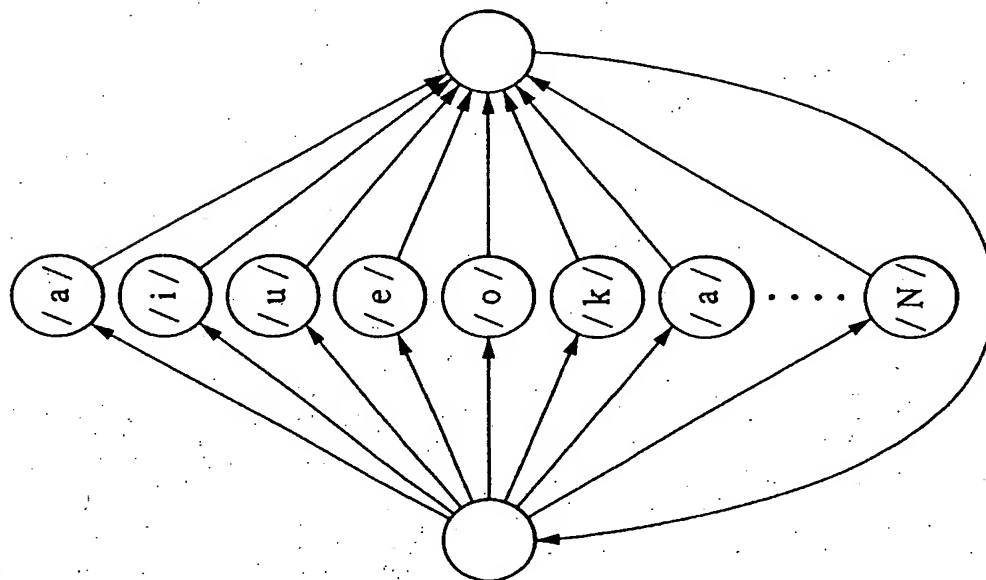


FIG.20B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP02/04378

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl⁷ G10L15/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl⁷ G10L15/00-17/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1926-1995	Toroku Jitsuyo Shinan Koho	1994-2002
Kokai Jitsuyo Shinan Koho	1971-2002	Jitsuyo Shinan Toroku Koho	1996-2002

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JICST FILE (JOIS)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 8-83092 A (Nippon Telegraph And Telephone Corp.), 26 March, 1996 (26.03.96), Full text; all drawings (Family: none)	1-17
A	JP 10-83195 A (Oki Electric Industry Co., Ltd.), 31 March, 1998 (31.03.98), Full text; all drawings (Family: none)	1-17
A	JP 9-65176 A (Fujitsu General Ltd.), 07 March, 1997 (07.03.97), Full text; all drawings (Family: none)	1-17

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

28 May, 2002 (28.05.02)

Date of mailing of the international search report

11 June, 2002 (11.06.02)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl¹ G10L15/00

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl¹ G10L15/00-17/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926~1995年
 日本国公開実用新案公報 1971~2002年
 日本国登録実用新案公報 1994~2002年
 日本国実用新案登録公報 1996~2002年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP 8-83092 A (日本電信電話株式会社) 1996.03.26 全文, 全図 (ファミリーなし)	1-17
A	JP 10-83195 A (沖電気工業株式会社) 1998.03.31 全文, 全図 (ファミリーなし)	1-17
A	JP 9-65176 A (株式会社富士通ゼネラル) 1997.03.07 全文, 全図 (ファミリーなし)	1-17

☐ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」 口頭による開示、使用、展示等に言及する文献
 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献
 「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」 同一パテントファミリー文献

国際調査を完了した日

28.05.02

国際調査報告の発送日

11.06.02

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)
 郵便番号100-8915
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

榎本 剛

5C 9379

電話番号 03-3581-1101 内線 3541



This Page is inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images
problems checked, please do not report the
problems to the IFW Image Problem Mailbox**